

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 19:06:11 ; Search time 380 Seconds
(without alignments)
268.307 Million cell updates/sec

Title: US-09-831-272-7
Perfect score: 24
Sequence: 1 cagccaccaagagaccagat 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1808:*
2: geneseqn1908:*
3: geneseqn2000:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	24	100.0	24	3	AAA27970 Box S eli
2	19.8	82.5	431	7	ABx62031 Arabidops
3	19.8	82.5	2868	6	ABZ12691 Arabidops
4	19.8	82.5	2868	7	ADA68110 Arabidops
5	19.8	82.5	141	3	AAA52645 Eosinoph
6	19.8	82.5	3062	6	ABK84368 Human
7	19.8	82.5	3062	6	ABN96885 Gene #338
8	19.8	82.5	3062	6	ABN96885 Gene #338
9	19.8	82.5	3062	6	ABN96885 Gene #338
10	19.8	82.5	3062	6	ABN96885 Gene #338
11	19.8	82.5	3062	6	ABN96885 Gene #338
12	19.8	82.5	3062	6	ABN96885 Gene #338
13	19.8	82.5	3062	6	ABN96885 Gene #338
14	19.8	82.5	3062	6	ABN96885 Gene #338
15	19.8	82.5	3062	6	ABN96885 Gene #338
16	19.8	82.5	3062	6	ABN96885 Gene #338
17	19.8	82.5	3062	6	ABN96885 Gene #338
18	19.8	82.5	3062	6	ABN96885 Gene #338
19	19.8	82.5	3062	6	ABN96885 Gene #338
20	19.8	82.5	3062	6	ABN96885 Gene #338
21	19.8	82.5	3062	6	ABN96885 Gene #338
22	19.8	82.5	3062	6	ABN96885 Gene #338
23	19.8	82.5	3062	6	ABN96885 Gene #338

24	18.2	75.8	3187	4	ABU05460 Drosophi
25	18.2	75.8	5159	9	ADD44763 Human gen
26	18.2	75.8	6353	4	ABU07173 Drosophi
27	18.2	75.8	8429	4	ABU07172 Drosophi
28	18.2	75.8	110000	3	AAE22303-4 Continuat
29	18.2	75.8	436	8	ACD06211 Human
30	17.8	74.2	446	8	ACD06210 Human
31	17.8	74.2	529	8	ACD06213 Human
32	17.8	74.2	110000	6	ABX90521-16 Continuat
33	17.6	73.3	693	7	ACA31733 Frokaryo
34	17.6	73.3	746	4	AAU01196 Human rep
35	17.6	73.3	746	4	ABU96655 Human tes
36	17.6	73.3	1398	5	AAE82726 DNA encod
37	17.6	73.3	1657	4	ABU18193 Drosophi
38	17.6	73.3	3117	5	AAE90916 DNA encod
39	17.6	73.3	5304	7	ADA98924 Human sec
40	17.6	73.3	5304	7	ADA44528 Human sec
41	17.6	73.3	5304	7	ADC20962 Human sec
42	17.6	73.3	7055	4	ABU18192 Drosophi
43	17.6	73.3	79467	8	ADA02717 Mouse Nfa
44	17.6	73.3	79467	9	ADB72455 Mouse Nfa
45	17.6	73.3	154465	6	ADD28763 Human AKA

ALIGNMENTS

RESULT 1	AAA27970 standard; DNA; 24 BP.
AAA27970	
15-AUG-2000 (first entry)	
Box S elicitor-responsive cis-element nucleotide sequence.	
Box S, elicitor-responsive cis-element; parsley; p1 promoter; chimeric promoter; pathogen infection; transgenic plant; resistance; herbicide; local response; genetic engineering; disease resistant crop; eli 7; ss.	
Petioselinum crispum.	
WO200029592-A2.	
25-MAY-2000.	
12-NOV-1999; 99WO-EP008710.	
12-NOV-1998; 98EP-00121160.	
27-NOV-1999; 99EP-00116981.	
(PLAC) MAX PLANCK GES FOERDERUNG.	
Kirsch C, Logemann R, Hahlbrock K, Rushton P, Somseich I; WPI; 2000-387804/33.	
Chimeric promoters mediating gene expression in plants upon pathogen infection, useful for transgenic plant production comprises at least one cis-acting element to direct elicitor-specific expression.	
Claim 1; Page 29; 73pp; English.	
This sequence represents Box S, an elicitor-responsive cis-element necessary for the expression of the Parsley eli 7 genes. The present invention relates to chimeric promoters capable of mediating local gene expression in plants upon pathogen infection. The chimeric promoters comprise at least one cis-element (see AAA27964-A27979) capable of directing elicitor-specific expression, and a minimal promoter. The chimeric promoters are useful for producing a transgenic plant which has attained resistance or improved resistance against a pathogen. The cis-	

CC acting element, chimeric promoter, recombinant gene encoding the chimeric
 CC promoter, vector comprising the chimeric promoter and a compound capable
 CC of activating the chimeric promoter are useful for producing pathogen
 CC resistant plants, and for identifying and/or producing compounds capable
 CC of conferring induced resistance to a pathogen in a plant. A compound
 CC which specifically activates or inhibits genes activated in a plant when
 CC attacked by a pathogen is also useful as a plant protective agent or a
 CC herbicide. The chimeric promoter provides rapid and local response to
 CC pathogen attack but shows negligible activity in uninfected parts of the
 CC plants and therefore is most suitable for the engineering of disease
 CC resistant crops

CC SQ Sequence 24 BP; 10 A; 8 C; 5 G; 1 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 3; Length 24;
 Beat Local Similarity 100.0%; Pred. No. 0.79;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGCCAGAAAT 24
 DB 1 CAGCCACCAAGAGAGCCAGAAAT 24

RESULT 2
 ABX62031/C
 ID ABX62031 standard; DNA; 431 BP.
 XX
 AC ABX62031;
 XX
 DT 25-FEB-2003 (first entry)
 XX
 DE Arabidopsis thaliana expressed sequence related polynucleotide #146.
 XX
 XX Transgenic plant; plant; genetically modified cell; environmental stress;
 KW ribozyme creation; disease resistance; stress tolerance;
 KW fungicide screening; insecticide screening; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN US2002040490-A1.
 XX
 PD 04-APR-2002.
 XX
 PF 26-JAN-2001; 2001US-00770423.
 XX
 PR 27-JAN-2000; 2000US-0178512P.
 XX

(GORL/) GORLACH J.
 PA (ANY/) AN Y.
 PA (HAM/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUY/) YU Y.
 PA (RAME/) RAMEAKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHEW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WORS/) WORSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALJE/) ALLEN K.
 PA (HOFF/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 XX

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 XX
 DR WPI; 2003-110411/10.

XX
 PT Novel Arabidopsis thaliana nucleic acid useful for identifying homologous
 PT or related genes, and to create genetically modified and transgenic
 PT organisms, such as plant cells and plants.

XX
 PS Claim 1; SEQ ID NO 146; 431p; English.

XX
 CC The invention describes an Arabidopsis thaliana nucleic acid (I). The
 CC polypeptide (II) encoded by (I), transgenic plant (III) or genetically
 CC modified cell (IV) are useful for screening a candidate agent for its
 CC biological effect, by combining the candidate agent with (II), (III) or
 CC (IV), and determining the effect of the candidate agent on (II), (III) or
 CC (IV). (I) is useful for identifying homologous or related genes, for
 CC producing compositions that modulate the expression or function of its
 CC encoded protein, for mapping functional regions of the protein, in
 CC diagnosis, for studying associated physiological pathways, for genetic
 CC manipulation of cells, preferably plant cells, in screening assays of
 CC various plant strains to determine the strains that are capable of
 CC withstanding a particular disease or environmental stress, for enhancing
 CC or inhibiting production of biosynthetic product in a plant, for
 CC producing polypeptides, as probes for the detection of mRNA in biological
 CC samples, to generate additional copies of (I), to generate ribozymes or
 CC oligonucleotides, as single stranded DNA probes or as triple-strand
 CC forming oligonucleotides, and to create genetically modified and
 CC transgenic organisms, such as plant cells and plants. (II) or (III) is
 CC useful for introducing or improving disease resistance and stress
 CC tolerance in plants, screening biological active agents, e.g.,
 CC fungicides, insecticides, etc., and for elucidating biochemical pathways.
 CC (III) is useful as crops for their enhanced disease resistance, enhanced
 CC traits of interest, for screening programs, as crops which exhibit
 CC enhanced tolerance to environmental stress, or to produce a factor. This
 CC sequence represents a nucleic acid that may correspond to naturally
 CC occurring Arabidopsis thaliana expressed sequences. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from the US patent office at
 CC seqdata.uspto.gov/sequence.html?docID=99909770423
 XX

CC SQ Sequence 431 BP; 124 A; 77 C; 103 G; 127 T; 0 U; 0 Other;

Query Match 82.5%; Score 19.8; DB 7; Length 431;
 Beat Local Similarity 91.3%; Pred. No. 63;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGCCAGAAAT 24
 DB 367 AACCCACCAAGAGAGCCAGAAAT 345

RESULT 3
 ABZ12691/C
 ID ABZ12691 standard; DNA; 2668 BP.
 XX
 AC ABZ12691;
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 496.
 XX
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200216655-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US026685.
 XX
 PR 24-AUG-2000; 2000US-0227866P.
 PR 26-JAN-2001; 2001US-0264647P.
 PR 22-JUN-2001; 2001US-0300111P.
 XX
 PA (SCRI) SCRIPPS RBS INST.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Harper JF, Kreps J, Wang X, Zhu T;
 PI WPI; 2002-304127/34.
 DR
 XX Identifying a stress condition to which a plant cell has been exposed and
 PT producing plants with increased tolerance to these abiotic stresses.
 XX
 PS Claim 144; SEQ ID NO 496; 577pp + Sequence Listing; English.
 CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 CC
 XX Sequence 2868 BP; 871 A; 581 C; 682 G; 734 T; 0 U; 0 Other;
 SQ
 Query Match 82.5%; Score 19.8; DB 6; Length 2868;
 Best Local Similarity 91.3%; Pred. No. 74;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 AGCCACCAAGAGAGACCCAGAAAT 24
 DB 694 AACCCACCAAGAGAGACCCAAAT 672
 RESULT 4
 ADA68110/c
 ID ADA68110 standard; DNA; 2868 BP.
 XX
 AC ADA68110;
 XX
 DT 20-NOV-2003 (first entry)
 DE Arabidopsis thaliana gene, SEQ ID 373.
 XX
 KM Plant; bacterial infection; fungal infection; viral infection; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001WO-IB001105.
 XX
 PR 22-JUN-2001; 2001WO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 PI WPI; 2003-175290/17.
 DR
 XX Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 PS Claim 6; SEQ ID NO 373; 899pp; English.
 CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 SQ Sequence 2868 BP; 871 A; 581 C; 682 G; 734 T; 0 U; 0 Other;
 XX
 Query Match 82.5%; Score 19.8; DB 7; Length 2868;
 Best Local Similarity 91.3%; Pred. No. 74;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 AGCCACCAAGAGAGACCCAGAAAT 24
 DB 694 AACCCACCAAGAGAGACCCAAAT 672
 RESULT 5
 AAA52645/c
 ID AAA52645 standard; DNA; 141 BP.
 XX
 AC AAA52645;
 XX
 DT 07-DEC-2000 (first entry)
 DE Eosinophil activating peptide gene 33.
 XX
 KM Eosinophil activation; human; allergy; eosinophilia; cancer;
 KM inflammation; signalling peptide; cell adhesion peptide;
 KM G-protein coupled receptor; secreted protein; cell surface protein; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200032630-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 06-DEC-1999; 99WO-US028773.
 XX
 PR 04-DEC-1998; 98US-0111006P.
 XX
 PA (SEAR) SEARLE & CO G D.
 XX
 PI Dotson SB, Ma X;
 PI WPI; 2000-465041/40.
 DR
 XX Novel nucleic acids derived from activated eosinophil cells useful for
 PT treating allergic diseases such as asthma comprises a specific nucleotide
 PT sequence.
 XX
 PS Claim 1; Page 114; 125pp; English.
 XX
 CC The present invention relates to a number of nucleotide sequences which
 CC encode proteins involved in the activation of eosinophils (AAA52601-
 CC A52679). Eosinophils are involved in immune reactions, and these genes
 CC and their proteins provide possible targets for new drugs to combat
 CC diseases such as asthma, allergic rhinitis, atopic dermatitis,
 CC anaphylaxis, allergic bronchopulmonary aspergillosis, eczema, psoriasis,
 CC emphysema, leukaemia, lymphomas, ovarian cancer, pneumonia, immune
 CC disorders, some connective tissue disorders, and inflammatory conditions
 CC including septic shock, arthritis, nephritis, inflammatory bowel disease
 CC and Crohn's disease
 XX
 SQ Sequence 141 BP; 17 A; 40 C; 49 G; 35 T; 0 U; 0 Other;
 XX
 Query Match 79.2%; Score 19; DB 3; Length 141;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAGCCACCAAGAGAGACC 19
 |||||
 DB 102 CAGCCACCAAGAGAGACC 84

RESULT 6
 ID ABRK4368 standard; cDNA; 3062 BP.
 AC ABRK4368;
 DT 14-AUG-2002 (first entry)
 DE Human cDNA differentially expressed in granulocytic cells #939.

Human; BS; granulocytic cell; DNA chip; bacterial infection;
 viral infection; parasitic infection; protozoal infection;
 fungal infection; sterile inflammatory disease; psoriasis;
 rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 cardiac reperfusion injury; renal reperfusion injury; ARDS;
 adult respiratory distress syndrome; inflammatory bowel disease;
 Crohn's disease; ulcerative colitis; periodontal disease;
 granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.
 PN WO200226999-A2.
 PD 11-APR-2002.
 PF 03-OCT-2001; 2001WO-US030821.
 PR 03-OCT-2000; 2000US-0237189P.
 PA (GENE-) GENE LOGIC INC.
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 DR WPI; 2002-435328/46.

PT Detecting granulocyte activation by detecting differential expression of
 genes associated with granulocyte activation, which serves as diagnostic
 markers that is useful for monitoring disease states and drug toxicity.

PS Claim 1; SEQ ID NO 939; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing the
 CC expression level to an expression level in an unactivated GC, where
 CC differential expression of Gs is indicative of GCA. Also included are
 CC modulating (M2) GA by contacting GC with an agent that alters the
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent
 CC capable of modulating GCA or an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease using the gene expression
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease, by detecting the level of
 CC expression in a sample of the tissue of gene(s) from Gs, where the level
 CC of expression of the gene is indicative of inflammation; (4) treating
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease, by contacting a tissue having inflammation with an
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
 CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
 CC for screening an agent capable of modulating GCA preferably in an
 CC inflammation in a tissue; M4 is useful for detecting an inflammation
 CC (especially chronic) in a tissue, an allergic response in a subject,
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's

CC disease, ulcerative colitis, periodontal disease; also bacterial
 CC infection, viral infection, parasitic infection, protozoal infection,
 CC fungal infection and MS is useful for treating one of the above
 CC conditions. The present sequence represents a gene differentially
 CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3062 BP; 547 A; 870 C; 1009 G; 636 T; 0 U; 0 Other;
 SQ

Query Match 79.2%; Score 19; DB 6; Length 3062;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAGCCACCAAGAGAGACC 19
 |||||
 DB 1081 CAGCCACCAAGAGAGACC 1063

RESULT 7
 ID ABR96885/C
 AC ABR96885;
 DT 13-AUG-2002 (first entry)
 DE Gene #3383 used to diagnose liver cancer.

Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 metastatic liver tumour; cytostatic; expression profile; disease state;
 disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.
 PN WO200229103-A2.
 PD 11-APR-2002.
 PF 02-OCT-2001; 2001WO-US030589.
 PR 02-OCT-2000; 2000US-0237054P.
 PA (GENE-) GENE LOGIC INC.
 PI Horne D, Alvares C, Pires-da-Silva S, Vockley JG;
 DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.

PS Claim 1; SEQ ID NO 3383; 298bp; English.

CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABR93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3062 BP; 547 A; 870 C; 1009 G; 636 T; 0 U; 0 Other;
 SQ

Query Match 79.2%; Score 19; DB 6; Length 3062;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
 |||||
 DB 1081 CAGCCACCAAGAGAGACC 1063

RESULT 8
 ADD18955/C
 ID ADD18955 standard; DNA; 3062 BP.

AC ADD18955;
 XX
 DT 15-JAN-2004 (first entry)

DE Human disease related protein DNA sequence SeqID444.

XX human; disease state; cytostatic; antiinflammatory; ophthalmological;
 XX antiarteriosclerotic; vulnery; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transportation; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing; gene; ds.

OS Homo sapiens.

PN WO2003018621-A2.

PD 06-MAR-2003.

XX 23-AUG-2002; 2002WO-GB003892.

PF 23-AUG-2001; 2001GB-00020558.

PR 05-OCT-2001; 2001GB-00024037.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

PA Kingman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

PI WPI; 2003-290046/28.

DR P-PSDB; ADD18954.

XX New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

PS Claim 27; SEQ ID NO 444; 424pp; English.

XX This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnery activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein encoding DNA sequence of the invention.

SQ Sequence 3062 BP; 547 A; 870 C; 1009 G; 636 T; 0 U; 0 Other;

Query Match 79.2%; Score 19; DB 9; Length 3062;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
 |||||
 DB 1081 CAGCCACCAAGAGAGACC 1063

RESULT 9
 ABZ68460/C
 ID ABZ68460 standard; DNA; 3205 BP.

AC ABZ68460;
 XX
 DT 22-APR-2003 (first entry)

DE Nucleotide sequence of human NOV1 polypeptide.

XX Human; NOV1; hexokinase 3; splice variant; metabolic disorder; diabetes;
 KW obesity; anorexia; cancer; metastatic melanoma; cancer; enzyme;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW immune disorder; systemic lupus erythematosus; haematopoietic disorder;
 KW cirrhosis; transplant; autoimmune disease; allergy; infection; vaccine;
 KW high blood pressure; haemophilia; wasting disorder; gene therapy; gene;
 KW ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 75..1148
 FT /*tag= a
 FT /product= "NOV1"

PN WO2003000918-A2.

PD 03-JAN-2003.

XX 21-JUN-2002; 2002WO-US019650.

PF 21-JUN-2001; 2001US-0299949P.

PR 22-JUN-2001; 2001US-0300290P.

PR 09-AUG-2001; 2001US-0311285P.

PR 05-OCT-2001; 2001US-0327345P.

PR 09-OCT-2001; 2001US-0327892P.

XX (CURA-) CURAGEN CORP.

PA Grose WM, Alsobrook JP, Lepley DM, Burgees CE, Bader JS;

PI Bansal A, Pena CE, Shinkets RA;

DR WPI; 2003-184053/18.

DR P-PSDB; ABP70634.

XX New NOV-1 polypeptides and polynucleotides, useful for manufacturing a

PT medicament for treating or preventing disorders or syndromes, including

PT metabolic disorders, cancer, neurodegenerative and autoimmune disorders.

XX Claim 21; Page 116; 147pp; English.

XX The present sequence encodes a polypeptide, designated NOV1. NOV1 is a
 CC member of the hexokinase 3-like family. It is an alternative splice
 CC variant. The NOV-1 polypeptide, nucleic acid or antibody is useful for
 CC manufacturing a medicament for treating or preventing disorders or
 CC syndromes, including metabolic disorders (e.g. diabetes, obesity, or
 CC anorexia), cancer (e.g. metastatic melanoma), cancer-associated
 CC disorders, neurodegenerative disorders (e.g. Alzheimer's disease,
 CC Parkinson's disease), immune disorders (e.g. systemic lupus
 CC erythematosus), haematopoietic disorders, cirrhosis, transplantation,
 CC autoimmune disease, allergies, viral/bacterial/parasitic infections, high
 CC blood pressure, haemophilia, and wasting disorders associated with
 CC chronic and various cancers. The polypeptides may be used as immunogen to
 CC produce antibodies, as vaccines, and to screen for potential agonist and
 CC antagonist compounds. The nucleic acids are also useful in gene therapy,
 CC in chromosome mapping, tissue typing, and in forensic identification of a

CC biological sample
XX
SQ Sequence 3205 BP; 571 A; 900 C; 1061 G; 673 T; 0 U; 0 Other;
Query Match 79.2%; Score 19; DB 7; Length 3205;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 CAGCCACCAAGAGAGACC 19
DB 1081 CAGCCACCAAGAGAGACC 1063
RESULT 10
AB268462/c
ID AB268462 standard; DNA; 3205 BP.
XX
AC AB268462;
XX
DT 22-APR-2003 (first entry)
XX
DE Nucleotide sequence of human NOV1 variant SNP1.
XX
KW Human; NOV1; hexokinase 3; splice variant; metabolic disorder; diabetes;
KW obesity; anorexia; cancer; metastatic melanoma; cancer; enzyme;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW immune disorder; systemic lupus erythematosus; haematopoietic disorder;
KW cirrhosis; transplant; autoimmune disease; allergy; infection; vaccine;
KW high blood pressure; hemophilia; wasting disorder; gene therapy;
KW single nucleotide polymorphism; SNP; ss.
XX
OS Homo sapiens.
XX
PN W02003000918-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019650.
XX
PR 21-JUN-2001; 2001US-0299949P.
PR 22-JUN-2001; 2001US-0300290P.
PR 09-AUG-2001; 2001US-0311285P.
PR 05-OCT-2001; 2001US-0327345P.
PR 09-OCT-2001; 2001US-0327892P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Grosse WM, Alsbrook JP, Lepley DM, Burgess CE, Bader JS;
PI Bansal A, Pena CE, Shinkets RA;
XX
DR WPI; 2003-184053/18.
XX
PT New NOV-1 polypeptides and polynucleotides, useful for manufacturing a
PT medicament for treating or preventing disorders or syndromes, including
PT metabolic disorders, cancer, neurodegenerative and autoimmune disorders.
XX
PS Claim 46; Page 121; 147pp; English.
XX
CC The present sequence represents a variant of NOV1, comprising a single
CC nucleotide polymorphism (SNP). NOV1 is a member of the hexokinase 3-like
CC family. It is an alternative splice variant. The NOV-1 polypeptide,
CC nucleic acid or antibody is useful for manufacturing a medicament for
CC treating or preventing disorders or syndromes, including metabolic
CC disorders (e.g. diabetes, obesity, or anorexia), cancer (e.g. metastatic
CC melanoma), cancer-associated disorders, neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), immune disorders (e.g.
CC systemic lupus erythematosus), haematopoietic disorders, cirrhosis,
CC transplantation, autoimmune disease, allergies, viral/bacterial/parasitic
CC infections, high blood pressure, haemophilia, and wasting disorders
CC associated with chronic and various cancers. The polypeptides may be used
CC as immunogen to produce antibodies, as vaccines, and to screen for
CC potential agonist and antagonist compounds. The nucleic acids are also
CC useful in gene therapy, in chromosome mapping, tissue typing, and in

CC forensic identification of a biological sample
XX
SQ Sequence 3205 BP; 571 A; 901 C; 1061 G; 672 T; 0 U; 0 Other;
Query Match 79.2%; Score 19; DB 7; Length 3205;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 CAGCCACCAAGAGAGACC 19
DB 1081 CAGCCACCAAGAGAGACC 1063
RESULT 11
ADCl0127/c
ID ADCl0127 standard; DNA; 3205 BP.
XX
AC ADCl0127;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human NOVX polypeptide coding sequence SEQ ID NO: 147.
XX
KW ds; gene; cytosolic; antidiabetic; anorectic; cerebroprotective;
KW neuroprotective; antiinflammatory; gene therapy; antisense therapy;
KW thymometric; NOVX; pathology; cancer; diabetes; obesity;
KW endocrine disorder; CNS disorder; inflammatory disorder;
KW chromosome mapping; tissue typing; predictive medicine.
XX
OS Homo sapiens.
XX
PN W02003000842-A2.
XX
PD 03-JAN-2003.
XX
PF 04-JUN-2002; 2002WO-US017443.
XX
PR 04-JUN-2001; 2001US-0295607P.
PR 04-JUN-2001; 2001US-0295612P.
PR 06-JUN-2001; 2001US-0296404P.
PR 06-JUN-2001; 2001US-0296418P.
PR 07-JUN-2001; 2001US-0296575P.
PR 11-JUN-2001; 2001US-0297414P.
PR 12-JUN-2001; 2001US-0295573P.
PR 12-JUN-2001; 2001US-0297567P.
PR 14-JUN-2001; 2001US-0298285P.
PR 15-JUN-2001; 2001US-0298528P.
PR 18-JUN-2001; 2001US-0299133P.
PR 19-JUN-2001; 2001US-0299230P.
PR 21-JUN-2001; 2001US-0299949P.
PR 22-JUN-2001; 2001US-0300177P.
PR 26-JUN-2001; 2001US-0300883P.
PR 28-JUN-2001; 2001US-0301530P.
PR 28-JUN-2001; 2001US-0301530P.
PR 28-JUN-2001; 2001US-0301550P.
PR 03-JUL-2001; 2001US-0302951P.
PR 31-JUL-2001; 2001US-0308890P.
PR 14-SEP-2001; 2001US-0322297P.
PR 25-SEP-2001; 2001US-0324669P.
PR 03-DEC-2001; 2001US-0337477P.
PR 14-DEC-2001; 2001US-0341562P.
PR 21-FEB-2002; 2002US-0358656P.
PR 21-FEB-2002; 2002US-0359122P.
PR 22-FEB-2002; 2002US-0358978P.
PR 22-FEB-2002; 2002US-0359034P.
PR 22-FEB-2002; 2002US-0359035P.
PR 22-FEB-2002; 2002US-0359121P.
PR 27-FEB-2002; 2002US-0359646P.
PR 01-MAR-2002; 2002US-0360858P.
PR 12-MAR-2002; 2002US-0363430P.
PR 12-MAR-2002; 2002US-0363676P.
PR 10-APR-2002; 2002US-0371346P.
PR 10-MAY-2002; 2002US-0379444P.
PR 04-JUN-2002; 2002US-00379444.

XX (CURA-) CURAGEN CORP.
PA Agsee M., Anderson DW, Berghs C, Casman SJ, Catterton E;
XX DiPippo VA, Edinger SR, Eisen A, Ellerman K, Gangolli EA;
PI Gerlach VL, Gorman L, Guo X, Herrmann JL, Hjalt T, Ji W, Kekuda R;
XX Khramtsov NV, Li L, Liu X, Malyankar UM, Miller CE, Miller I;
PI Ort T, Padigaru M, Paturajan M, Pena CE, Rastelli L, Rieger DK;
PI Rothenberg ME, Shenoy SG, Shinkets RA, Smltson G, Spaderna SK;
PI Spytek KA, Stone DJ, Vernet CAM, Zhong H, Zhong W, Alsobrook JP;
PI Burgess CE, Lepley DM;
XX MPI; 2003-210149/20.
DR P-PSDB; ADC10128.
XX
XX New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
XX Claim 20; SEQ ID NO 147; 772pp; English.
XX
XX The invention relates to novel isolated polypeptides, mature form of the
CC polypeptide, a sequence that is 95% identical to the polypeptide or the
CC polypeptide comprising one or more conservative substitutions. The NOVX
CC polypeptide is useful for treating or preventing a pathology associated
CC with the polypeptide e.g. disorders associated with aberrant expression
CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC endocrine, CNS and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. This sequence corresponds to the coding sequence
CC for one of the polypeptides of the invention.
XX
SQ Sequence 3205 BP; 571 A; 900 C; 1061 G; 673 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19; DB 9; Length 3205;
Best Local Similarity 100.0%; Pred. NO. 1.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGAGGCC 19
DB 1081 CAGCCACCAAGAGAGGCC 1063
XX
RESULT 12
AAV00090/c
ID AAV00090 standard; cDNA; 3692 BP.
XX
AC AAV00090;
XX
DT 17-MAR-1998 (first entry)
XX
DE Rat hexokinase III encoding cDNA.
XX
KM Yeast; trehalose-6-phosphate synthase; hexokinase; ribozyme; stimulator;
KM insulin; glucose; diabetes; ss.
XX
OS Rattus sp.
XX
XX Key Location/Qualifiers
FH 81..2855
FT CDS /*tag= a
FT FT /product= "Hexokinase_III"
XX
XX MO9726357-A1.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-US000787.
XX
XX 19-JAN-1996; 96US-00588983.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
PA

PA (BETA-) BETAGENE INC.
XX
XX Newgard CB, Han H, Thispen AE, Normington KD;
PI
XX MPI; 1997-385343/35.
XX
DR P-PSDB; AAM37430.
XX
XX Mammalian cell with reduced activity of low Km hexokinase - caused by
PT specific ribozyme or stimulator of trehalose-6-phosphate production,
PT particularly for in vitro or in vivo insulin production.
XX
XX Disclosure; Page 195-198; 265pp; English.
XX
XX A mammalian cell has been developed comprising an effective amount of a
CC low Km hexokinase-inhibitor selected from: (a) an agent that stimulates
CC production of trehalose-6-phosphate (T6P); or (b) a ribozyme specific for
CC low Km hexokinase. The present sequence encodes rat hexokinase III. The
CC cells are particularly used to produce insulin, in response to glucose or
CC other secretagogues, either in vitro or in vivo (for treating diabetes),
CC but may also be used to produce many other therapeutic proteins, e.g.
CC from a gene under control of the insulin promoter which is therefore
CC responsive to glucose. Growth-inhibited cells (having altered hexokinase
CC activity) are used to produce proteins, e.g. insulin or antibodies, in
CC vivo or in vitro. Reduction in low Km hexokinase activity provides cells
CC in which insulin secretion is induced at glucose concentrations closer to
CC the normal range than in the parent cell (nearly homeostatic secretion).
CC Implanted cells of reduced low Km hexokinase activity are expected to
CC survive longer in the host
XX
SQ Sequence 3692 BP; 821 A; 969 C; 1112 G; 790 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19; DB 2; Length 3692;
Best Local Similarity 100.0%; Pred. NO. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGAGGCC 19
DB 1087 CAGCCACCAAGAGAGGCC 1069
XX
RESULT 13
AAV00160/c
ID AAV00160 standard; cDNA; 3692 BP.
XX
AC AAV00160;
XX
DT 17-MAR-1998 (first entry)
XX
DE Rat hexokinase III encoding cDNA.
XX
KM Hexokinase; competitive enzyme; inhibitor; insulin; diabetes; B cell;
KM glucose; glucokinase; ss.
XX
OS Rattus sp.
XX
XX Key Location/Qualifiers
FH 81..2855
FT CDS /*tag= a
FT FT /product= "Hexokinase_III"
XX
XX MO9726322-A2.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-US000786.
XX
XX 19-JAN-1996; 96US-00588976.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX (BETA-) BETAGENE INC.
XX
XX (UNMS) UNIV MICHIGAN STATE.
PI Newgard CB, Han H, Becker TC, Wilson JB;

```
XX MPI: 1997-385327/35.
DR P-PSDB; AAM37443.
XX Mammalian cell with reduced activity of low Km hexokinase - caused by
PT competitive enzyme inhibitor, particularly an insulin secreting cell for
PT in vitro or in vivo insulin production.
XX
PS Disclosure: Page 195-198; 254pp; English.
XX
CC A mammalian cell has been developed which includes enough of a
CC competitive inhibitor of low Km (Michaelis constant) hexokinase to reduce
CC the hexokinase activity in the cell. The present sequence encodes rat
CC hexokinase III. The cells are particularly used to produce insulin, in
CC response to glucose or other secretagogues, either in vitro or in vivo
CC (for treating diabetes), but may also be used to produce many other
CC therapeutic proteins, e.g. from a gene under control of the insulin
CC promoter and thus responsive to glucose. Inhibition of the inhibitor can
CC be used more generally to reduce the growth rate of cells, e.g.
CC hydridomas that produce antibodies
XX
SQ Sequence 3692 BP; 821 A; 969 C; 1112 G; 790 T; 0 U; 0 Other;
Query Match 79.2%; Score 19; DB 2; Length 3692;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGAGACC 19
DB 1087 CAGCCACCAAGAGAGACC 1069

RESULT 14
ABT42065/C
ID ABT42065 standard; DNA; 3692 BP.
XX
AC ABT42065;
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 1767.
XX
KM Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KM database; drug screening; toxicity assay; rat; ds.
XX
OS Rattus norvegicus.
XX
PN WO200295000-A2.
XX
PD 28-NOV-2002.
XX
PE 22-MAY-2002; 2002MO-US016173.
XX
PR 22-MAY-2001; 2001US-0292335P.
PR 13-JUN-2001; 2001US-0297523P.
PR 19-JUN-2001; 2001US-0298925P.
PR 10-JUL-2001; 2001US-0303807P.
PR 10-JUL-2001; 2001US-0303808P.
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 27-SEP-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 06-DEC-2001; 2001US-0336144P.
PR 19-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.
PR 21-FEB-2002; 2002US-0357844P.
PR 15-MAR-2002; 2002US-0364134P.
PR 08-APR-2002; 2002US-0370144P.
PR 08-APR-2002; 2002US-0370206P.
PR 08-APR-2002; 2002US-0370247P.
```

```
PR 17-APR-2002; 2002US-0372794P.
PR 21-APR-2002; 2002US-0371679P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX MPI: 2003-148464/14.
XX
DR Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
PS Example 4; Page; 446pp; English.
XX
CC The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the
CC progression of a toxic effect of a compound, predicting the renal
CC toxicity of a compound, or identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 3692 BP; 821 A; 969 C; 1112 G; 790 T; 0 U; 0 Other;
Query Match 79.2%; Score 19; DB 7; Length 3692;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGAGACC 19
DB 1087 CAGCCACCAAGAGAGACC 1069

RESULT 15
ADB58419/C
ID ADB58419 standard; DNA; 3692 BP.
XX
AC ADB58419;
XX
DT 04-DEC-2003 (first entry)
XX
DE Toxicity-related gene, SEQ ID 3445.
XX
KM Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KM drug screening; toxicity assay; ds.
XX
OS Unidentified.
XX
PN WO2003064624-A2.
XX
PD 07-AUG-2003.
XX
PE 31-JAN-2003; 2003MO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
```


XX
DR WPI: 2003-689530/65.

XX
PT Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.

PS
XX Claim 1; SEQ ID NO 3445; 1156bp; English.

CC The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 3692 BP; 821 A; 969 C; 1112 G; 790 T; 0 U; 0 Other;

Query Match 79.2%; Score 19; DB 9; Length 3692;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAGCCACCAAGAGACCC 19
|||
DB 1087 CAGCCACCAAGAGACCC 1069

Search completed: July 30, 2004, 20:33:04
Job time : 386 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 19:11:31 ; Search time 2503 Seconds

(without alignments)
415.594 Million cell updates/sec

Title: US-09-831-272-7

Perfect score: 24

Sequence: 1 cagccaccaagagagaccagacat 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_in.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vtl.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	24	AX025607	AX025607 Sequence
2	24	100.0	3549	AF239835	AF239835 Petroselin
3	19.8	82.5	2868	AX050801	AX050801 Sequence
4	19.8	82.5	2868	AX651563	AX651563 Sequence
5	19.8	82.5	2962	AY133782	AY133782 Arabidops
6	19.8	82.5	3107	AF081202	AF081202 Arabidops
7	19.8	82.5	3295	AF088801	AF088801 Lilium lo
8	19.8	82.5	3383	AF080601	AF080601 Arabidops
9	19.8	82.5	81843	AL672309	AL672309 Mouse DNA
10	19.8	82.5	91071	AC002339	AC002339 Arabidops
11	19.4	80.8	813	BV071807	BV071807 S212P6764
12	19.4	80.8	173763	AC134552	AC134552 Mus muscu
13	19.4	80.8	183349	AC087416	AC087416 Mus muscu
14	19.4	80.8	213625	AL805918	AL805918 Mouse DNA
15	19.2	80.0	1603	VCCYNA	X69193 V. carduncul
16	19.2	80.0	130709	AF082029	AF082029 Hemerocal
17	19.2	80.0	189891	AC102845	AC102845 Mus muscu
18	19.2	80.0	191132	AC119058	AC119058 Papio anu
19	19.2	80.0	191132	AC126895	AC126895 Rattus no
20	19.2	80.0	192082	AL591490	AL591490 Mouse DNA
21	19.2	80.0	206145	BX813306	BX813306 Mus muscu
22	19.2	80.0	214393	BX537302	BX537302 Mouse DNA
23	19.2	80.0	226299	AC112320	AC112320 Rattus no
24	19.2	80.0	247220	AC072807	AC072807 Mus muscu
25	19.2	80.0	255871	AC095418	AC095418 Rattus no
26	19.2	80.0	260367	AC099427	AC099427 Rattus no
27	19.2	79.2	3062	AX410736	AX410736 Sequence
28	19.2	79.2	3062	HSU51333	U51333 Human hexok
29	19.2	79.2	3073	BC028129	BC028129 Homo sapi
30	19.2	79.2	3592	AR068840	AR068840 Sequence
31	19.2	79.2	3692	AR070140	AR070140 Sequence
32	19.2	79.2	3692	RNU73859	U73859 Rattus norv
33	19.2	79.2	121282	AC027318	AC027318 Homo sapi
34	19.2	79.2	166534	AC034209	AC034209 Homo sapi
35	19.2	79.2	168168	AC012283	AC012283 Homo sapi
36	19.2	79.2	220152	AC119372	AC119372 Rattus no
37	19.2	79.2	221407	AC139592	AC139592 Rattus no
38	19.2	79.2	267520	AC095489	AC095489 Rattus no
39	18.8	78.3	400	GI7829	GI7829 human STS
40	18.8	78.3	2228	AY069309	AY069309 Drosophila
41	18.8	78.3	2452	BC044275	BC044275 Xenopus 1
42	18.8	78.3	2999	AY128932	AY128932 Arabidops
43	18.8	78.3	3335	AF081203	AF081203 Arabidops
44	18.8	78.3	3548	AY093052	AY093052 Arabidops
45	18.8	78.3	73181	AC117455	AC117455 Homo sapi

ALIGNMENTS

RESULT 1
AX025607
LOCUS AX025607 24 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 7 from Patent WO0029592.
ACCESSION AX025607
VERSION AX025607.1 GI:10187275
KEYWORDS
SOURCE
ORGANISM
Petroselinum crispum (parsley)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Apiales; Apiaceae; Apioidae; apioid
superclade; Apium clade; Petroselinum.
REFERENCE
1 Logemann, E., Somsch, I., Hahlbrock, K., Kirsch, C. and Rushton, P.

TITLE Chimeric promoters capable of mediating gene expression in plants
JOURNAL upon pathogen infection and uses thereof
Patent: WO 0029592-A 7 25-MAY-2000;
MAX PLANCK GESELLSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSSICH IMRE
(DE) ; HAHNBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL
(GB)

FEATURES
SOURCE location/Qualifiers
1..24
/organism="Petroselinum crispum"
/mol_type="unassigned DNA"
/db_xref="taxon:4043"

ORIGIN
Query Match 100.0%; Score 24; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAGCCACCAAGAGAGAGCCAGAAAT 24
1 CAGCCACCAAGAGAGAGCCAGAAAT 24

Db 1 CAGCCACCAAGAGAGAGCCAGAAAT 24

RESULT 2 3549 bp DNA linear PLN 22-OCT-2000
AF239835
LOCUS Petroselinum crispum fatty acid desaturase/hydroxylase-like protein
DEFINITION EL17.1 (EL17.1) gene, complete cds; and fatty acid
desaturase/hydroxylase-like protein EL17.2 (EL17.2) gene, partial
cds.
AF239835
AF239835.1 GI:10945374
Petroselinum crispum (parsley)
Petroselinum crispum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Apiales; Apiaceae; Apioideae; aploid
superclade; Apium clade; Petroselinum.
1 (bases 1 to 3549)
Kirsch,C., Takamiya-Wik,M., Schmelzer,E., Hahnbrock,K. and
Somssich,I.E.
A novel regulatory element involved in rapid activation of parsley
EL17 gene family members by fungal elicitor or pathogen infection
Mol. Plant Pathol. 1 (4), 243-251 (2000)
2 (bases 1 to 3549)
Kirsch,C., Takamiya-Wik,M. and Somssich,I.E.
Direct Submision
Submitted (29-FEB-2000) Biochemistry, Max-Planck-Institut fuer
Zuechtungsforschung, Carl-von-Linne Weg 10, Koeln, NRW D-50829,
Germany

FEATURES
SOURCE location/Qualifiers
1..3549
/organism="Petroselinum crispum"
/mol_type="genomic DNA"
/db_xref="taxon:4043"
/note="from cultured cells"
380..1782
/gene="EL17.1"
380..1782
/gene="EL17.1"
/product="fatty acid desaturase/hydroxylase-like protein
EL17.1"
551..1702
/gene="EL17.1"
/codon_start=1
/product="fatty acid desaturase/hydroxylase-like protein
EL17.1"
/protein_id="AAG24521.1"
/db_xref="GI:10945375"

gene
mRNA
CDS

REFERENCE
JOURNAL
AUTHORS
TITLE
JOURNAL
TITLE
JOURNAL

TERRKOIIVSDVAILAVIYGLYRLYLKGFAMVFCYGGPLLVNMMFTLLIINH
HPSVPPYDSFEMDLRGALCTYRDGILINKVPHNVCNAHVCHIFSMIPHYGLEAT
EAKKPLIGDTYYQDGTPIITAMTRBKKECTYKDESGTGYWYRKDT"
3218..>3549
/gene="EL17.2"
3218..>3549
/gene="EL17.2"
/product="fatty acid desaturase/hydroxylase-like protein
EL17.2"
3384..>3549
/gene="EL17.2"
/codon_start=1
/product="fatty acid desaturase/hydroxylase-like protein
EL17.2"
/protein_id="AAG24522.1"
/db_xref="GI:10945376"
/translation="MGAGRMSAPSNGKTEBALRRAPHEKPPPTIGDLKAI PAHC
FKSLVTSFRY"

ORIGIN
Query Match 100.0%; Score 24; DB 8; Length 3549;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAGCCACCAAGAGAGAGCCAGAAAT 24
1 CAGCCACCAAGAGAGAGCCAGAAAT 24

Db 1 CAGCCACCAAGAGAGAGCCAGAAAT 24

RESULT 3 2868 bp DNA linear PAT 27-SEP-2002
AX505801
LOCUS Arabidopsis thaliana (thale cress)
DEFINITION Sequence 496 from Patent WO0216655.
ACCESSION AX505801
VERSION AX505801.1 GI:23387038
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1
Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
Stress-regulated genes of use
same, and methods of use
Patent: WO 0216655-A 496 28-SEP-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)

FEATURES
SOURCE location/Qualifiers
1..2868
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN
Query Match 82.5%; Score 19.8; DB 6; Length 2868;
Best Local Similarity 91.3%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AGCCACCAAGAGAGAGCCAGAAAT 24
2 AGCCACCAAGAGAGAGCCAGAAAT 24

Db 694 AACACCAAGAGAGAGCCAGAAAT 672

RESULT 4 2868 bp DNA linear PAT 22-MAR-2003
AX651563
LOCUS Arabidopsis thaliana
DEFINITION Sequence 373 from Patent WO03000898.
ACCESSION AX651563
VERSION AX651563.1 GI:29154381
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE

1. A. S. Gerasimov, *Otchekhovye i krasnye tsvety*, Vol. V, Moscow-Leningrad, Gosizdatkhi, 1960.

REFERENCE
AUTHORS
TITLE
JOURNAL

1
Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Karatgiri, F., Qian, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
Plant genes involved in defense against pathogens
Parent: WO 03000898-A 373 03-JAN-2003;
Parent: WO 03000898-A 373 03-JAN-2003;
Parent: WO 03000898-A 373 03-JAN-2003;

```

FEATURES
source      Location/Qualifiers
1. .2868

```

ORIGIN

Query Match Similarity	82.5%	Score	19.8	DB	6	length	2868		
Best Local Similarity	91.3%	Pred.	No. 52						
Matches	21	Conservative	0	Mismatches	2	Indels	0	Gaps	0

```
QY      2 AGCCACCAAGAGGACCAGAAT 24
          | ||||| ||||| |||
Db      694 AACCAACAAGAGGACCCAAAT 672
```

RESULT 5	2962 bp	mRNA	linear	PLN 18-SEP-2007
AY133782/c				
LOCUS				
DEFINITION				
	AY133782	2962 bp	mRNA	linear
	Arabidopsis thaliana clone U11765	putative villin 2 protein		
	(AC2941740)	mnNA, complete cds.		

ABSTRACTS	FBI-CHINA.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana

REFERENCE
Yamada, K., Chan, M. M., Chang, C. H., Dale, J. M., Deng, J. M., Hsuan, V. W.
1 (baae 1 to 2362)

TITLE	Arabidopsis Open Reading Frame (ORF) Clones
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2962)
AUTHORS	Yamada, K., Chan, M. M., Chang, C. H., Dale, J. M., Deng, J. M., Hsuan, V. W.,

JOURNAL COMMENT
Submitted (17-JUN-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pUN1 (ORF) clones using the RAL1 cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriunt, M., Wallender, E.R., Wong, C., Wu, H.C., Yu, G., Yuan, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Miranda, M., Nguyen, M., and Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Eckert, J.R. and

Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES
source

3'UTR
ORIGIN

Query Match	82.5%	Score 19.8;	DB 8;	Length 2962;
Best Local Similarity	91.3%	Pred. No. 52;		
Matches	21;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0

QY	2	AGCCACCAAGAGGAGCCCAAGAT	24
Db	694	AACCAACCAAGAGGAGCCCAAAAT	672

AF081202/c	3107 bp	mRNA	linear	PLN 07-FEB-2000
LOCUS	AF081202			
DEFINITION	Arabidopsis thaliana villin 2 (VLN2) mRNA, complete cds.			

KEYWORDS	.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana

REFERENCE
AUTHORS
TITLE

Bakeriyea; Vitidiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 3107)
Klaine, O., Friederich, E., Kost, B., Louvard, D. and Chua, N. H.
Villin-like actin-binding proteins are expressed ubiquitously in
Arabidopsis

JOURNAL Plant Physiol. 122 (1), 35-48 (2000)
 MEDLINE 20098711
 PUBLISHED 10631247
 REFERENCE 2 (bases 1 to 3107)
 AUTHORS Klahre, U. and Chua, N.-H.
 TITLE Direct Submission
 JOURNAL Submitted (31-JUL-1998) Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

FEATURES
 source
 Location/Qualifiers
 1..3107
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /cultivar="Landsberg"
 /db_xref="taxon:3702"
 /chromosome="II"
 1..3107
 /gene="VILN2"
 1..2931
 /note="VILN2"
 /note="Atvillin; similar to protein encoded by TILN.7.16 gene"
 /codon_start=1
 /product="villin 2"
 /protein_id="AAC31606.1"
 /db_xref="GI:3415115"
 /translation="MSTKVLDPAGAGQKPKTRIMRIENFPAVPVPSKSEHKRYMGD
 TYVLTQTTONGGAVLPDIHFWIGKDTSDPAGTAIVTVLDAVIGRAVQHRVIG
 HESDKFSYFKPCITPLRGVVSQKTYEEVEFTRILYTCGRKRAIRLQVPPARSL
 NHDVFLIDTBEKTYIQFGANSNIQERAKALEVVQYLKDKYBGTCTVALVDGKLDT
 BSDGAEFWLPGFAPIGKRVAGNDIDYESTPRLYCTIDGKMEPIDGDSKSMKL
 TKCYLDCGAEIYVWGVTQVDRKAQSAEFLSENPRTKATHTVRVIOGESIS
 KSNPDSWPSGATPGNEBGRKVAALKQGVGKIGAKAPVNEIDPLPLESGKL
 EAVYVNGVTKPLPKEDIKYSGDCYIVLTYHSGEKDEYPLSCWRGKSLPEDD
 TAILANTMSILKRGPVQGRITYGKEKPPQVALPQPMVTLKGLSGSYKSMSESS
 TDETYTPESIALVOVSGVHNNAVQETVALSINSECFLLQSGTSMFLMHQNSST
 HSEDLATKVAEFLKPGITLLKRAKGETSSTFWPALGKQNFSTKASSETIRPDLF
 SFAPNRGFEVBEIYNFADDLTEDIYLDTHAEVFWGQCYEPEKQVFIQK
 YIDLAGSLGHPVPIYIKINEGPCFTTFSPMDATKATVOGNSPCKKSLFEGH
 HVYDKSNGMOGIRORBAALANSAFNSNSNPAYSDORLNEHSDGQRPORALAL
 ALSSAFNSSSSTKSPPPRPVGTSGASORAAVAALSOIVLAKKSPDTSPPRRST
 SSNPADIPLTAEADBEASVAGLEAESEVSPADETAKQETEGQSGSEIQPG
 APTVEQLRAKSENPVTGIDPKRREAVILSEEPQSVFGIEKAEAFNNLPRWKDILLKKK
 FDLF"

ORIGIN
 Query Match 82.5%; Score 19.8; DB 8; Length 3107;
 Best Local Similarity 91.3%; Pred. No. 52;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGACCAGAAAT 24
 |||||
 Db 694 AACCCAAAGAGACCAGAAAT 672

RESULT 7
 AF088901/c 3295 bp mRNA linear PLN 22-FEB-2000
 LOCUS Lilium longiflorum actin bundling protein ABP135 (Y5-7) mRNA,
 DEFINITION complete cds.
 ACCESSION AF088901
 VERSION AF088901.1 GI:5880463
 KEYWORDS
 SOURCE
 ORGANISM
 .
 Lilium longiflorum (trumpet lily)
 Lilium longiflorum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;
 Lilium.
 1 (bases 1 to 3295)
 REFERENCE
 AUTHORS Vidal, L., Yokota, E., Cheung, A., Shimmen, T. and Hepler, P. K.
 TITLE The 135 kDa actin-bundling protein from Lilium longiflorum pollen
 JOURNAL Protoplasma 209, 283-291 (1999)
 REFERENCE 2 (bases 1 to 3295)

AUTHORS Vidal, L., Yokota, E., Cheung, A., Wu, H. M., Shimmen, T. and
 Hepler, P. K.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-1998) Biology, University of Massachusetts,
 Morrill Science Center, Amherst, MA 01003, USA

FEATURES
 source
 Location/Qualifiers
 1..3295
 /organism="Lilium longiflorum"
 /mol_type="mRNA"
 /cultivar="White Lily"
 /db_xref="taxon:4690"
 /tissue_type="pollen grains"
 /note="Isolated with antibodies raised against F-actin
 binding activity purified from pollen grains"
 1..3295
 /gene="Y5-7"
 123..3020
 /note="Y5-7"
 /note="similar to Arabidopsis thaliana villin protein, animal and fungal gelosolin and villins; demonstrates bundling activity; may be post-translationally modified"
 /codon_start=1
 /product="actin bundling protein ABP135"
 /protein_id="AAD54660.1"
 /db_xref="GI:5880464"
 /translation="MANSSKNLDPAGVGORIGTETIRIENFQVELPKSDHGFYS
 GDSYIVLTQTTONGGAVLPDIHFWIGKDTSDPAGTAIVTVLDAVIGRAVQHRVIG
 OCHESDFFLSYFKPCITPLRGVVSQKTYEEVEFTRILYTCGRKRAIRLQVPPARSL
 SLNHDVFLIDTBEKTYIQFGANSNIQERAKALEVVQYLKDKYBGTCTVALVDGKLDT
 AASGSEFWLPGFAPIGKRVAGNDIDYESTPRLYCTIDGKMEPIDGDSKSMKL
 ENNKCYLDCGAEIYVWGVTQVDRKAQSAEFLSENPRTKATHTVRVIOGESIS
 KSNPDSWPSGATPGNEBGRKVAALKQGVGKIGAKAPVNEIDPLPLESGKL
 EAVYVNGVTKPLPKEDIKYSGDCYIVLTYHSGEKDEYPLSCWRGKSLPEDD
 TAILANTMSILKRGPVQGRITYGKEKPPQVALPQPMVTLKGLSGSYKSMSESS
 TDETYTPESIALVOVSGVHNNAVQETVALSINSECFLLQSGTSMFLMHQNSST
 HSEDLATKVAEFLKPGITLLKRAKGETSSTFWPALGKQNFSTKASSETIRPDLF
 SFAPNRGFEVBEIYNFADDLTEDIYLDTHAEVFWGQCYEPEKQVFIQK
 YIDLAGSLGHPVPIYIKINEGPCFTTFSPMDATKATVOGNSPCKKSLFEGH
 HVYDKSNGMOGIRORBAALANSAFNSNSNPAYSDORLNEHSDGQRPORALAL
 ALSSAFNSSSSTKSPPPRPVGTSGASORAAVAALSOIVLAKKSPDTSPPRRST
 SSNPADIPLTAEADBEASVAGLEAESEVSPADETAKQETEGQSGSEIQPG
 APTVEQLRAKSENPVTGIDPKRREAVILSEEPQSVFGIEKAEAFNNLPRWKDILLKKK
 FDLF"

ORIGIN
 Query Match 82.5%; Score 19.8; DB 8; Length 3295;
 Best Local Similarity 91.3%; Pred. No. 52;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGACCAGAAAT 24
 |||||
 Db 822 AGCCACCAAGAGACCAGAAAT 800

RESULT 8
 AY080601/c 3383 bp mRNA linear PLN 18-SEP-2002
 LOCUS Arabidopsis thaliana putative villin 2 protein (At2g41740) mRNA,
 DEFINITION complete cds.
 ACCESSION AY080601
 VERSION AY080601.1 GI:19310557
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
 1 (bases 1 to 3383)
 REFERENCE
 AUTHORS Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, D., Chung, M. K.,
 Goldsmith, A. D., Lee, J. M., Quach, H. L., Toriumi, M., Yu, G., Bowser, L.,
 Carminci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
 Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
 Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,
 Satou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K.,

RESULT 10	LOCUS	AC002339/c	91071 bp	DNA	linear	PLN 11-MAR-2002
DEFINITION	Arabidopsis thaliana chromosome 2 clone T11A7 map C1C11C08, complete sequence.	AC002339				
ACCESSION	AC002339					
VERSION	AC002339.3					
KEYWORDS	HTG.					
SOURCE	Arabidopsis thaliana (thale cress)					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Buckwold, J., Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.					
AUTHORS	1 (bases 1 to 91071) Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C., Spriggs, T.A., Mason, T.M., Kellavage, A.R., Adams, M.D., Somerville, C.R. and Venter, J.C.					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 91071) Lin, X.					
AUTHORS	Direct Submission					
TITLE	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712					
JOURNAL	Medical Center Dr., Rockville, MD 20850, USA					
REFERENCE	3 (bases 1 to 91071) Town, C.D. and Kaul, S.					
AUTHORS	Direct Submission					
TITLE	Submitted (27-FEB-2002) The Institute for Genomic Research, 9712					
JOURNAL	Medical Center Dr., Rockville, MD 20850, USA, cdtown@igr.org					
COMMENT	On Apr 18, 2002 this sequence version replaced gi:6598350.					
FEATURES	Location/Qualifiers					
source	1..91071					
	/organism="Arabidopsis thaliana"					
	/mol_type="genomic DNA"					
	/cultivar="Columbia"					
	/db_xref="taxon:3702"					
	/chromosome="2"					
	/map="C1C11C08"					
	/clone="T11A7"					
	1..13047					
	/note="overlap with BAC clone T6D20					
	(U90439.82778..95824)."					
	184..2478					
	/gene="At2g41890"					
	/note="synonym: T11A7.1"					
	<184..2478					
	/gene="At2g41890"					
	184..2478					
	/gene="At2g41890"					
	/codon_start=1					
	/product="putative receptor-like protein kinase"					
	/protein_id="AA114823.1"					
	/db_xref="gi:20196890"					
	/translat="MKMLRALICSLVFLPAPQIVSEIQSLVVGENTLVVNNNN					
	GDPAFGFNPGLNRFSITQIMPNNIS.PYQKRYVWAGVGVNDSVSEFLTRNGK					
	ELVLPDLSGLVPMVSKTRPSVSGALRDGNGVTLADREIIVSGSGCTPTLLPR					
	QKPAFEMLRPAASNSRSYISLHEDSGRLKLRNHSITTMVSGNVEYKKKKKRN					
	GAVLISGALFLIEDQDLRPVWSVFGEDHNTVFRFLRLDRDNLNRYSWNEBSRLW					
	KPVWVAENOCRFATCGSOVCSFNSGCTCNCPEFNAFVSDPKCLVPYQKGCCKS					
	GFNMAKFNLELYGVIYPPANDSVISQISSQRCCKLSENSACTAVTYNDGPOCRMKL					
	TRYISGSDPSLSSISYKTCIDPLAVPNVSKSPVTVKSHSICPLVAGATSTTGK					
	LVLFLPGLQIVVYVRRKKLAKKARFPAKATPKGVMIPSYDEIQAAMDNDNNIN					
	GPQIKGVMPNEMELVAVKEVATLTERKFRSSNAKICTMHNKLANLEGCISGLRR					
	FLVETVAKNGSLDHTVDPLNSKLTWRIRTDCLSVKALCYLHMECREVSHGNLNL					
	CGNLLIGLDEAKLTVEYFGCLCAADKVEDGKTVLALITGRYEBGVSSWVREWM					
	GGKRTVVDKGLGCGFVDEBELERVLIRISFCVQTDRLRPSMGVEVVKLETSLLSP					
	PPPPFCASSPPTNSSESSQSLYEP"					
	810..833					
	/rpt_family="GAA)n"					
	complement(3322..5857)					
	/gene="At2g41890"					
	/note="synonym: T11A7.23; supported by cDNA:					
	repeat_region					
	gene					

[illegible]


```

/gene="At2g41850"
/note="synonym: T11A7.5"
join(<16477>.16731,16830,.16961,17046,.17216,17458,.17665,
17823,.17904,17999,.18107,18206,.18319,18409,.>18618)
/gene="At2g41850"
join(16477,.16731,16830,.16961,17046,.17216,17458,.17665,
17823,.17904,17999,.18107,18206,.18319,18409,.18618)
/gene="At2g41850"
/codon_start=1
/product="putative polysaccharonase"
/protein_id="AAC02763.1"
/db_xref="GI:2335094"
/translation="MARCTNLVTFELMALLPFSWCKASRIENPNYDHSYKPKFSDSL
IKRREDITGLRSEFVRASLRTPTTVASVSDGANGDGTDDTOAFVAMKKAQSSNGAVN
LVRPKNTYLLKSLIQTGPCNSILTVQIGTLLSASGRSDYDI SKMIWFDVNINISV
DGGDPTVADGEGTMMQNSCKRNKAKALFENYSKLIIVNKLKRNAGQIOISIEKSN
VQVSNVTVAPADSPRTDGIHITNONTIVSESLIIGTGDGISIESGSONVOINDITC
GPHGHSIGSLGSDNSKAFVSGVTVGAKLSTGDSNVRKTTQGGSGTSNIIIPQIQ
MNVKNKPIIIDQYCDKSKCTEKSEAVQVKNVYRDISGTSASENALTPNCKNYPQQ
GIYLDVNIKGGATCTNNANVVDKGALVPOCNST"
/complement(17300,.17333)
/rpt_family="A)n"
/rpt_family="18742
18645,.18742
18997,.20710
18997,.20710"
/gene="At2g41840"
/note="synonym: T11A7.6; supported by full length cDNA:
Cereb1.17415"
join(18997,.19833,20359,.20710)
/gene="At2g41840"
join(19138,.19833,20359,.20520)
/gene="At2g41840"
/codon_start=1
/product="40S ribosomal protein S2"
/protein_id="AAC02764.1"
/db_xref="GI:2335095"
/translation="MAERGGERGVERGEGDFGRFGGRGGRGDRGGRGGRGR
GGVASEEKKVPYTKGLVLAAGHIKQIEQIYLSLPVKEYQIIDMIIGTLLDEYWK
IMVQKOTRAGROTRFPAFVGVGDGNGHVLGVKSGSEVATLRGMIILAKLSVVR
RGTWGNKIKGPRHVPCKYKNGKSGSVYRWVAPRSGIYVAPRPKRVLPAGIDVFT
SSRSGKTLGNFVKATIFDCLQTYGFLTFEWFKEPFSNSPYQEHDFLASKLSTSK
PDVVEDQA"
/gene="At2g41830"
24083,.30105
/note="synonym: T11A7.7; predicted by geneScan; similar to
GP_2191175 AF007270"
join(<24083>.24229,24605,.24679,24862,.24986,25189,.25303,
25475,.25558,25660,.25707,25758,.25871,26076,.26241,
26473,.26905,27234,.27749,27973,.28122,28266,.28341,
28553,.28680,28855,.29049,29166,.29303,29391,.29501,
29600,.29839,29941,.>30105)
/gene="At2g41830"
join(24083,.24229,24605,.24679,24862,.24986,25189,.25303,
25475,.25558,25660,.25707,25758,.25871,26076,.26241,
26473,.26905,27234,.27749,27973,.28122,28266,.28341,
28553,.28680,28855,.29049,29166,.29303,29391,.29501,
29600,.29839,29941,.30105)
/gene="At2g41830"

Query Match      82.5%; Score 19.8; DB 8; Length 91071;
Best Local Similarity 91.3%; Pred. No. 57;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY      2 AGCCACCAAAAGAGACCCAGAAAT 24
          |||||
Db      64909 AACCAACAAAGAGACCCAAAT 64887

RESULT 11
BV0071807      813 bp      DNA      linear      STS 31-MAY-2003
LOCUS      5212P676F6E6.T0 CZBCH11/E1 Mus musculus STS genomic, sequence
DEFINITION      tagged site.

```

ACCESSION	BV071807
VERSION	BV071807.1 GI:31187602
KEYWORDS	STS.
SOURCE	Mus musculus
ORGANISM	Mus musculus (house mouse)
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 813) Wade,C.E.M., Kulbokas,E.J., III, Kirby,A.W., Zody,M.C., Mullikin,J.C., lander,E.S., Lindblad-Toh,K. and Daly,M.J. The mosaic structure of variation in the laboratory mouse genome Nature 420 (6915), 574-578 (2002)
TITLE	Nature 420 (6915), 574-578 (2002)
JOURNAL	
MEDLINE	22354684
PUBMED	12466852
COMMENT	Contact: Kerstin Lindblad-Toh Whitehead Institute for Biomedical Research, Center for Genome Research 320 Charles Street, Cambridge, MA 02141, USA Tel: 6175221477 Fax: 6172580903 Email: kersli@genome.wi.mit.edu Primer A: None Primer B: None STS size: 813 Protocol: WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SvimJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGCv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STS and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%. Location/Qualifiers 1..813 /organism="Mus musculus" /mol type="genomic DNA" /strain="CZECHII/EI" /db_xref="taxon:10090" /map="x" 8 26-777 5648606-5649357" /clone_id="CZECHII/EI" <1..>813
FEATURES	
source	
ORIGIN	
STS	
Query Match	80.8%; Score 19.4; DB 11; Length 813;
Best Local Similarity	95.2%; Pred. No. 80;
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
CY	1 CAGCCACCAGAAGGAGCCAG 21
DB	565 CAGCCAGCAAAAGAGGCCAG 585
RESULT 12	
LOCUS	AC134552 173763 bp DNA linear HTG 20-MAR-2003
DEFINITION	Mus musculus chromosome UNK clone Rf24-279K10, WORKING DRAFT
ACCESSION	AC134552
VERSION	AC134552.3 GI:29126544
KEYWORDS	HTG, HTGS, PHASE1, HTGS_DRAFT, HTGS_FULUTOP.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 173763) McPherson,J.D. and Waterson,R.H.
JOURNAL	The sequence of Mus musculus clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 173763) McPherson,J.D. and Waterson,R.H.
TITLE	Direct Submission

JOURNAL Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 173763)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (20-MAR-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Mar 20, 2003 this sequence version replaced gi:29124268.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Contact: submissions@watson.wustl.edu

Project Information -----

Center project name: M_BB0279K10

----- Summary Statistics -----

Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big_Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 172608 bases at least Q40

Consensus quality: 173012 bases at least Q30

Consensus quality: 173235 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2809: contig of 2809 bp in length

* 2810 2909: gap of unknown length

* 2910 26146: contig of 2337 bp in length

* 26147 86246: gap of unknown length

* 26247 89830: contig of 63584 bp in length

* 89831 89930: gap of unknown length

* 89931 173763: contig of 83833 bp in length.

Location/Qualifiers

1. 173763

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="X"

/clone="RP24-279K10"

1. 2809

/note="assembly_name:Contig23"

2910. 26146

/note="assembly_name:Contig24"

26247. 89830

/note="assembly_name:Contig25"

89931. 173763

/note="assembly_name:Contig26"

ORIGIN

Query Match 80.8%; Score 19.4; DB 2; Length 173763;

Best Local Similarity 95.2%; Pred. No. 93;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGACCCAG 21

|||||

Db 91787 CAGCCAGCAAGAGACCCAG 91807

|||||

RESULT 13

AC087416 183349 bp DNA linear ROD 15-NOV-2002

LOCUS AC087416

DEFINITION Mus musculus chromosome 2 clone rp23-224116 strain C57BL/6J, complete sequence.

ACCESSION AC087416 GI:25013342

VERSION AC087416.18

KEYWORDS HMG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 183349)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.

TITLE 1 (bases 1 to 183349)

JOURNAL Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Mus musculus BAC Clone rp23-224116

REFERENCE 2 (bases 1 to 183349)

AUTHORS Unpublished

TITLE Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

JOURNAL Submitted (01-JUN-2001) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 3 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2001) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 4 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (12-JUL-2001) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 5 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (14-JUL-2001) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 6 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 7 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 8 (bases 1 to 183349)

AUTHORS Kim,J., Shauli,S., Yao,Z. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT On Nov 15, 2002 this sequence version replaced gi:22002184.

----- Genome Center -----

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

Location/Qualifiers

1. 183349

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/chromosome="X"

/clone="rp23-224116"

/clone_lib="RPC1 - 23 Female (C57BL/6J) Mouse BAC Library"

ORIGIN

Query Match 80.8%; Score 19.4; DB 10; Length 183349;

Best Local Similarity 95.2%; Pred. No. 93;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY      2  AGCCACCAAGAGAGACCAGCA  22
Db      97066  AGCCACCAAGAGAGACTCACA  97086

RESULT 14
AL805918      213625 bp      DNA      linear      ROD 04-OCT-2002
LOCUS      AL805918
DEFINITION      Mouse DNA sequence from clone RP23-191L7 on chromosome 2, complete
sequence.
ACCESSION      AL805918
VERSION      AL805918.5
KEYWORDS      GI:233395461
SOURCE      HTG.
ORGANISM      Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 213625)
Leongamornlert, D.
Direct Submission
Submitted (04-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
humquerry@sanger.ac.uk
On Oct 1, 2002 this sequence version replaced GI:23337476.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web file: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.; EMBL; SW.;
SWISSPROT; Tr.; TrEMBL; Wp.; WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
RP23-191L7 is
from the RPCI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES
source
1..213625
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-191L7"
/clone_11b="RPCI-23"
ORIGIN
Query Match 80.8%; Score 19.4; DB 10; Length 213625;
Best Local Similarity 95.2%; Pced No. 93;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0
OY      2  AGCCACCAAGAGAGACCAGCA  22
Db      136743  AGCCACCAAGAGAGACTCACA  136763

```

```

LOCUS       VCCYNA/c                      1603 bp      mRNA      linear      PLN 03-OCT-1994
DEFINITION   v cardunculus mRNA for cynarase.
ACCESSION    X69193
VERSION      X69193.1 GI:499015
KEYWORDS     aspartic proteinase; cynarase.
SOURCE       Cynara cardunculus
ORGANISM     Cynara cardunculus
              Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              asterids; campanulids; Asterales; Asteraceae; Cardioideae;
              Cardueae; Cynara.
REFERENCE    1 (bases 1 to 1600)
AUTHORS      Cordeiro,M.C., Xue,Z.T., Pietrzak,M., Pais,M.S. and Brodelius,P.E.
TITLE        Isolation and characterization of a cDNA from flowers of Cynara
             cardunculus encoding cynrosin (an aspartic proteinase) and its use
             to study the organ-specific expression of cynrosin
JOURNAL      Plant Mol. Biol. 24 (5), 733-741 (1994)
MEDLINE      94250836
PUBMED       8193298
REFERENCE    2 (bases 1 to 1600)
AUTHORS      Brodelius,P.
TITLE        Direct Submision
JOURNAL      Submitted (13-NOV-1992) P. Brodelius, University of Lund, Dept of
             Plant Biochemistry, POB 7007, 220 07 Lund, SWEDEN
FEATURES     Location/Qualifiers
             source          1..1603
                        /organism="Cynara cardunculus"
                        /mol_type="mRNA"
                        /db_xref="taxon:4265"
                        /clone="Cypro 1s"
                        /c_lone="flower"
                        /clone_id="1g1l1-CDNA"
                        /dev_stage="floral development"
                        <1..1427
                        /codon_start=3
                        /product="cyprosin"
                        /protein_id="CAA8939.1"
                        /db_xref="GI:509163"
                        /db_xref="GOA:P40782"
                        /translation="LKKRKNILNHPGESHAGSNDANARRKYGVGNFRSDGELIALK
                        NYMAOAYGEIGETIPROKFTVLFDTSNLMWPSKCYFSAVLKATKEGITFLAKFGILGG
                        KNGKSAAIITYGTSTISGFSPDSVKLGILLIVRODPIEATKEGITFLAKFGILGG
                        LGFEISVGDAVPWMTMLNQGLVQEVREFFMYLRNADBDDEGEIVNGCVDPHFKEE
                        HTYPVTYQKGWQEGMDVLIGDKTTFCSASGCALADSGLSLAGTTITVTQNQAI
                        GAAVWSQQCKSLVDGYCKSMIEMLSEBPBKICQMUKCSFGSHDTSMILEAVD
                        KSKSSGSLHEVCCTCOMAMVMNONIQNETENTINYVDKLCERLPSPMGESADV
                        CSSLSMNTIAFTWGKTPTNLSPROYVALKVEGATTAICISGFTAMDVAIPHGMILIG
                        DFMGQHTVTDVGNLVRFAEA"
                        102..1424
                        /product="cyprosin"
                        /note="putative active site Asp-Thr /Ser-Gly"
                        207..215
                        /note="putative active site Asp-Thr /Ser-Gly"
                        768..776
                        /note="putative active site Asp-Thr /Ser-Gly"
                        1506..1511
                        1589
ORIGIN
Query Match      80.0%; Score 19.2; DB 8; Length 1603;
Best Local Similarity 87.5%; Pred. NO. 1e+02;
Matches         21; Mismatches      3; Indels      0; Gaps      0;

```

Miss Margaret Brown (Mrs. C. J. Brown)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 30, 2004, 20:19:46 ; Search time 70 Seconds
(without alignments)
190.269 Million cell updates/sec

Title: US-09-831-272-7

Perfect score: 24

Sequence: 1 cagccaccaagagagaccagaat 24

Scoring table: IDENTITY NUC

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/5B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	79.2	3692	2	US-08-588-983-17
2	19	79.2	3692	2	US-08-588-976-17
3	17.4	72.5	2606	2	US-08-701-233B-11
4	16.6	69.2	289	4	US-09-443-184-111
5	16.6	69.2	349	4	US-09-833-181-1287
6	16.6	69.2	2543	4	US-09-976-594-575
7	16.6	69.2	2636	4	US-09-976-594-383
8	16.2	67.5	375	4	US-09-702-705-961
9	16.2	67.5	375	4	US-09-702-705-1158
10	16.2	67.5	375	4	US-09-736-457-961
11	16.2	67.5	375	4	US-09-736-457-1158
12	16.2	67.5	375	4	US-09-614-124B-961
13	16.2	67.5	375	4	US-09-614-124B-1158
14	16.2	67.5	375	4	US-09-671-325-961
15	16.2	67.5	375	4	US-09-671-325-1158
16	16.2	67.5	571	4	US-09-736-457-921
17	16.2	67.5	571	4	US-09-736-457-921
18	16.2	67.5	571	4	US-09-614-124B-921
19	16.2	67.5	791	4	US-09-671-325-921
20	16.2	67.5	791	4	US-09-702-705-1805
21	16.2	67.5	791	4	US-09-736-457-1805
22	16.2	67.5	791	4	US-09-671-325-1805
23	16.2	67.5	1833	1	US-08-454-720A-36
24	16.2	67.5	3183	1	US-08-943-731-183
25	16.2	67.5	20084	3	US-08-943-731-5
26	16	66.7	889	4	US-09-227-357-88
27	16	66.7	1665	4	US-08-812-008-31

28	16	66.7	3839	4	US-09-245-248B-54	Sequence 54, Appl
29	16	66.7	29485	4	US-09-785-381-6	Sequence 6, Appl
30	15.8	65.8	109	2	US-08-427-097-8	Sequence 8, Appl
31	15.8	65.8	109	2	US-08-878-957-8	Sequence 8, Appl
32	15.8	65.8	399	4	US-09-188-930-63	Sequence 63, Appl
33	15.8	65.8	399	4	US-09-312-283C-63	Sequence 63, Appl
34	15.8	65.8	1695	2	US-08-878-957-31	Sequence 31, Appl
35	15.8	65.8	1746	2	US-08-427-097-29	Sequence 29, Appl
36	15.8	65.8	1746	2	US-08-878-957-29	Sequence 29, Appl
37	15.8	65.8	1752	1	US-08-427-097-13	Sequence 13, Appl
38	15.8	65.8	1752	1	US-08-427-097-13	Sequence 13, Appl
39	15.8	65.8	1752	1	US-08-427-097-27	Sequence 27, Appl
40	15.8	65.8	1752	2	US-08-878-957-13	Sequence 13, Appl
41	15.8	65.8	1752	2	US-08-878-957-19	Sequence 19, Appl
42	15.8	65.8	1752	2	US-08-878-957-27	Sequence 27, Appl
43	15.8	65.8	9103	1	US-08-466-033-182	Sequence 182, App
44	15.8	65.8	9103	1	US-08-444-733-182	Sequence 182, App
45	15.8	65.8	9103	2	US-08-464-134-182	Sequence 182, App

ALIGNMENTS

RESULT 1
US-08-588-983-17/c
Sequence 17, Application US/08588983
Patent No. 5854067
GENERAL INFORMATION:
APPLICANT: Christopher B. Newgard, et al.
TITLE OF INVENTION: Methods and Compositions
TITLE OF INVENTION: for Inhibiting Hexokinase
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,983
FILING DATE: Concurrently herewith
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Fussey, Shelley P.M.
REGISTRATION NUMBER: 39,458
REFERENCE/DOCKET NUMBER: UTSD:424/FUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
TELEX: n/a
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 3692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-588-983-17
Query Match 79.2%; Score 19; DB 2; Length 3692;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
Db 1087 CAGCCACCAAGAGAGACC 1069

RESULT 2
US-08-588-976-17/C
Sequence 17, Application US/08588976
Patent No. 5891717
GENERAL INFORMATION:
APPLICANT: Christopher B. Newgard, et al.
TITLE OF INVENTION: Methods and Compositions for
TITLE OF INVENTION: Inhibiting Hexokinase
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,976
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fusey, Shelley P.M.
REGISTRATION NUMBER: 39,458
REFERENCE/DOCKET NUMBER: UTSD:481/FUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
TELEX: n/a
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 3692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-588-976-17
Query Match 79.2%; Score 19; DB 2; Length 3692;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGACCC 19
Db 1087 CAGCCACCAAGAGACCC 1069
RESULT 3
US-08-701-233B-1/C
Sequence 1, Application US/08701233B
Patent No. 5861308
GENERAL INFORMATION:
APPLICANT: Preundschuh, Michael; Renner, Christoph
TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS ASSOCIATED WITH T CELL
TITLE OF INVENTION: ACTIVATION AND USES THEREOF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,233B
FILING DATE: 21-August-1996

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, NO. 5861308man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5440
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2606 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-701-233B-1
Query Match 72.5%; Score 17.4; DB 2; Length 2606;
Best Local Similarity 94.7%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 CCACCAAGAGACCCAGA 22
Db 1917 CCACCAAGAGACCCAGA 1899
RESULT 4
US-09-443-184-111/C
Sequence 11, Application US/09443184A
Patent No. 6372431
GENERAL INFORMATION:
APPLICANT: Cunningham, Mary Jane
APPLICANT: Zweigler, Gary
APPLICANT: Kaser, Matthew R.
APPLICANT: Panzer, Scott
APPLICANT: Selhammer, Jeffrey J.
APPLICANT: Yue, Henry
APPLICANT: Baughn, Mariah
APPLICANT: Azimzai, Valda
APPLICANT: Lal, Preeti
TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
FILE REFERENCE: PC-0007 US
CURRENT APPLICATION NUMBER: US/09/443,184A
CURRENT FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 138
SOFTWARE: PERL Program
SEQ ID NO 111
LENGTH: 289
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURES:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID NO. 6372431 700067537H1
US-09-443-184-111
Query Match 69.2%; Score 16.6; DB 4; Length 289;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CAGCCACCAAGAGACCCAGAA 23
Db 267 CAGCCACCAAGAGACCCAGAA 245
RESULT 5
US-09-833-381-1287/C
Sequence 1287, Application US/09833381
Patent No. 6672186
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
FILE REFERENCE: 5800-119
CURRENT APPLICATION NUMBER: US/09/833,381
CURRENT FILING DATE: 2001-04-11

```

; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1287
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(349)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1287
```

```

Query Match          69.2%; Score 16.6; DB 4; Length 349;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 CAGCCACCAAGAGAGCCAGAA 23
Db 45 CAGACACCAAGAGAGCCAGTA 23
```

```

RESULT 6
US-09-976-594-575/c
; Sequence 575, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 575
; LENGTH: 2543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1859458CH1
US-09-976-594-575
```

```

Query Match          69.2%; Score 16.6; DB 4; Length 2543;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 CAGCCACCAAGAGAGCCAGAA 23
Db 2081 CAGACACCAAGAGAGCCAGTA 2059
```

```

RESULT 7
US-09-976-594-383
; Sequence 383, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 383
; LENGTH: 2636
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 238026.4
US-09-976-594-383
```

```

Query Match          69.2%; Score 16.6; DB 4; Length 2636;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 CAGCCACCAAGAGAGCCAGAA 23
Db 906 CAGACACCAAGAGAGCCAGTA 928
```

```

RESULT 8
US-09-702-705-961/c
; Sequence 961, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 961
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-961
```

```

Query Match          67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 AGCCACCAAGAGAGCCAGAA 22
Db 163 AGCCACCAAGAGAGCCAGAA 143
```

```

RESULT 9
US-09-702-705-1158
; Sequence 1158, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
```

/ SEQ ID NO 1158
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-702-705-1158

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
DB 213 AGCCACCAAGAGAGACCAGA 233

RESULT 10
US-09-736-457-961/c
/ Sequence 961, Application US/09736457
/ Patent No. 6509448
/ GENERAL INFORMATION:

/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Fan, Liqun
/ APPLICANT: Wang, Aijun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C15
/ CURRENT APPLICATION NUMBER: US/09/736,457
/ CURRENT FILING DATE: 2000-12-13
/ NUMBER OF SEQ ID NOS: 1864
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 961
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-736-457-961

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
DB 163 AGCCACCAAGAGAGACCAGA 143

RESULT 11
US-09-736-457-1158
/ Sequence 1158, Application US/09736457
/ Patent No. 6509448
/ GENERAL INFORMATION:

/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Fan, Liqun
/ APPLICANT: Wang, Aijun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C15
/ CURRENT APPLICATION NUMBER: US/09/736,457
/ CURRENT FILING DATE: 2000-12-13
/ NUMBER OF SEQ ID NOS: 1864

/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1158
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-736-457-1158

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
DB 213 AGCCACCAAGAGAGACCAGA 233

RESULT 12
US-09-614-124B-961/c
/ Sequence 961, Application US/09614124B
/ Patent No. 6630574
/ GENERAL INFORMATION:

/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Fan, Liqun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C9
/ CURRENT APPLICATION NUMBER: US/09/614,124B
/ CURRENT FILING DATE: 2001-07-11
/ NUMBER OF SEQ ID NOS: 1668
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 961
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-614-124B-961

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
DB 163 AGCCACCAAGAGAGACCAGA 143

RESULT 13
US-09-614-124B-1158
/ Sequence 1158, Application US/09614124B
/ Patent No. 6630574
/ GENERAL INFORMATION:

/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C9
/ CURRENT APPLICATION NUMBER: US/09/614,124B
/ CURRENT FILING DATE: 2001-07-11
/ NUMBER OF SEQ ID NOS: 1668
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1158
/ LENGTH: 375

TYPE: DNA
ORGANISM: Homo sapien
US-09-614-124B-1158

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
Db 213 AGCCACCAAGAGAGACCAGA 233

RESULT 14
US-09-671-325-961/c
Sequence 961, Application US/09671325
Patent No. 6667154

GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedrick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C12
CURRENT APPLICATION NUMBER: US/09/671,325
CURRENT FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 1825
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 961
LENGTH: 375
TYPE: DNA
ORGANISM: Homo sapien
US-09-671-325-961

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
Db 163 AGCCACCAAGAGAGACCAGA 143

RESULT 15
US-09-671-325-1158
Sequence 1158, Application US/09671325
Patent No. 6667154

GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedrick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C12
CURRENT APPLICATION NUMBER: US/09/671,325
CURRENT FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 1825
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1158
LENGTH: 375
TYPE: DNA

ORGANISM: Homo sapien
US-09-671-325-1158

Query Match 67.5%; Score 16.2; DB 4; Length 375;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGAGACCAGA 22
|||||
Db 213 AGCCACCAAGAGAGACCAGA 233

Search completed: July 30, 2004, 22:06:01
JOD time : 72 secs

It's Buga Book (uspo)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 20:26:46 ; Search time 406 Seconds
(without alignments)
289.841 Million cell updates/sec

Title: US-09-831-272-7

Sequence: 1 cagccaccaagaggaccagaat 24

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA.*
1: /cgn2_6/pcodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/pcodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/pcodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/pcodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/pcodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/pcodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/pcodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/pcodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/pcodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/pcodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/pcodata/2/pubpna/US09_NEW_PUB.seq.*
12: /cgn2_6/pcodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/pcodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/pcodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/pcodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/pcodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/pcodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/pcodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/pcodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.8	82.5	431	US-09-770-423-146	Sequence 146, App
C 2	19.8	82.5	1451	US-10-424-599-65132	Sequence 65132, A
C 3	19.8	82.5	2868	US-09-938-842A-496	Sequence 496, App
C 4	19.8	82.5	2868	US-09-938-842A-496	Sequence 496, App
C 5	19.8	79.2	141	US-10-350-923B-45	Sequence 45, App1
C 6	19.8	79.2	3062	US-09-880-107-3382	Sequence 3382, App
C 7	19.8	79.2	3062	US-10-942-887-728	Sequence 728, App
C 8	19.8	79.2	3062	US-10-172-118-728	Sequence 728, App
C 9	19.8	79.2	3062	US-10-172-118-728	Sequence 728, App
C 10	19.8	79.2	3205	US-10-177-809-1	Sequence 1, App1
C 11	19.8	79.2	3205	US-10-177-809-1	Sequence 1, App1
C 12	19.8	79.2	3205	US-10-177-809-1	Sequence 3, App1
C 13	19.8	79.2	3692	US-10-152-319A-1767	Sequence 1767, App
C 14	19.8	79.2	3692	US-10-152-319A-1767	Sequence 1767, App

C 15	19	79.2	18772	US-09-764-861-63	Sequence 63, App1
C 16	19	79.2	18772	US-09-764-861-63	Sequence 63, App1
C 17	19	79.2	18772	US-10-115-928-63	Sequence 63, App1
C 18	18.2	75.8	625	US-10-027-632-211873	Sequence 211873, App1
C 19	18.2	75.8	625	US-10-027-632-211873	Sequence 211873, App1
C 20	18.2	75.8	625	US-10-027-632-211873	Sequence 211873, App1
C 21	18.2	75.8	625	US-10-027-632-211873	Sequence 211873, App1
C 22	18.2	75.8	2689	US-10-260-238-474	Sequence 474, App
C 23	18.2	75.8	3515	US-10-437-963-91409	Sequence 91409, App
C 24	18.2	75.8	36211	US-10-085-117-49	Sequence 49, App1
C 25	17.8	74.2	41	US-10-035-833A-123	Sequence 123, App
C 26	17.8	74.2	436	US-10-035-833A-5462	Sequence 5462, App
C 27	17.8	74.2	436	US-10-236-392-85	Sequence 85, App1
C 28	17.8	74.2	529	US-10-236-392-89	Sequence 89, App1
C 29	17.8	74.2	529	US-10-236-392-89	Sequence 89, App1
C 30	17.6	73.3	255	US-09-733-627A-2711	Sequence 2711, App
C 31	17.6	73.3	567	US-10-027-632-92107	Sequence 92107, App
C 32	17.6	73.3	567	US-10-027-632-92107	Sequence 92107, App
C 33	17.6	73.3	567	US-10-027-632-92107	Sequence 92107, App
C 34	17.6	73.3	567	US-10-027-632-92107	Sequence 92107, App
C 35	17.6	73.3	624	US-10-027-632-187229	Sequence 187229, App
C 36	17.6	73.3	624	US-10-027-632-187229	Sequence 187229, App
C 37	17.6	73.3	693	US-10-282-122A-19603	Sequence 19603, App
C 38	17.6	73.3	746	US-09-764-891-1197	Sequence 1197, App
C 39	17.6	73.3	79467	US-10-052-482-223	Sequence 223, App
C 40	17.6	73.3	162025	US-09-834-700-13	Sequence 13, App1
C 41	17.6	73.3	162025	US-09-834-700-14	Sequence 14, App1
C 42	17.6	73.3	162025	US-09-834-700-17	Sequence 17, App1
C 43	17.6	73.3	162025	US-09-834-700-18	Sequence 18, App1
C 44	17.6	73.3	162025	US-10-272-665-35	Sequence 35, App1
C 45	17.6	73.3	162025	US-10-272-665-36	Sequence 36, App1

ALIGNMENTS

RESULT 1
US-09-770-423-146/C
Sequence 146, Application US/09770423
Publication No. US20020040490A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jörn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Moessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krickler, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2028 (PAPA-01/PRV)
CURRENT APPLICATION NUMBER: US/09/770,423
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,512
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 146
LENGTH: 431
TYPE: DNA
ORGANISM: Arabidopsis thaliana

US-09-770-423-146

Query Match	82.5%	Score 19.8;	DB 13;	Length 431;
Best Local Similarity	91.3%	Pred. No. 38;		
Matches 21; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	2	AGCCACCAAGAGGAGCCAGAT	24
Db	367	AACCAACCAAGAGGAGCCAAAT	345

RESULT 2
US-10-424-599-65132

```

1      sequence 65132, Application US/10424559
2      Publication No. US20040031072A1
3      GENERAL INFORMATION:
4      APPLICANT: La Rosa Thomas J
5      APPLICANT: Kovalic David K
6      APPLICANT: Zhou Yihua
7      APPLICANT: Cao Yongwei
8      TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
9      TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
10     FILE REFERENCE: 38-21(53223)B
11     CURRENT APPLICATION NUMBER: US/10/424,599
12     CURRENT FILING DATE: 2003-04-28
13     NUMBER OF SEQ ID NOS: 285684
14     SEQ ID NO 65132
15     LENGTH: 1451
16     TYPE: DNA
17     ORGANISM: Glycine max
18     FEATURE:
19     OTHER INFORMATION: Clone ID: PAT_MRT3847_29827C.1
20     US-10-424-599-65132

```

Query Match	82.5%	Score 19.8;	DB 13;	Length 1451;
Best Local Similarity	91.3%;	Pred. No. 37;		
Matches	21; Conservative	0; Mismatches	2; Indels	0; Gaps 0;

QY	2	AGCCACCAAGAGGAGCCAGAAT	24
Db	427	AACCAACCAAGAGGAGCCCAAAAT	449

RESULT 3
US-09-93

```

: Sequence 496: Application US/09938842A
: Patent No. US2002160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT300-3
: CURRENT APPLICATION NUMBER: US/09/938,842A
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300,111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 496
: LENGTH: 2866
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-938-842A-496

```

Query Match	82.5%	Score 19.8;	DB 9;	Length 2868;
Best Local Similarity	91.3%;	Pred. No. 37;		
Matches	21; Conservative	0; Mismatches	2; Indels	0

```

QY      2 AGCCACCAAGAGGACCAGAAT 24
          | | | | | | | | | |
Db      694 AACCAACCAAGAGGACCCAAAT 672

```

RESULT 4

```

US-09-938-842A-436/C
; Sequence 496, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krebs, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 496
; LENGTH: 2668
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-496

```

Query Match	82.5%	Score 19.8	DB 11	Length 2868
Best Local Similarity	91.3%	Pred. No. 37		
Matches 21	Conservative 0	Mismatches 2	Indels 0	Gaps 0

OY	2	AGCCACCAAGAGACCCAGAAT	24
Db	694	AAACCAACCAAGAGACCCAAAT	672

RESULT 5
US-10-35

```

: Sequence 45 Application US/10350923B
: Publication No. US20040002084A1
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Ma, Xiao Jun
: TITLE OF INVENTION: Nucleic acid, polypeptides, vectors, and cells derived from actinophiles
: TITLE OF INVENTION: eosinophil cells
: FILE REFERENCE: S03157-01
: CURRENT APPLICATION NUMBER: US/10/350,923B
: CURRENT FILING DATE: 2003-01-24
: PRIOR APPLICATION NUMBER: US/ 09/454,280
: PRIOR FILING DATE: 1999-12-03
: PRIOR APPLICATION NUMBER: PCT/US99/28773
: PRIOR FILING DATE: 1999-06-12
: PRIOR APPLICATION NUMBER: US 60/111,006
: PRIOR FILING DATE: 1998-12-04
: NUMBER OF SEQ ID NOS: 79
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 45
: LENGTH: 141
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-350-923B-45

```

Query Match	79.2%	Score 19;	DB 16;	Length 141;
Best Local Similarity	100.0%	Pred. No. 89;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CAGCCACCAAGAGACCC	19	

QY 1 CAGCCACCAAGAGGACCC 19

Db 102 CAGCCACCAAGAGACC 84

RESULT 6
US-09-880-107-3382/c
; Sequence 3382, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Schert, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-MO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR FILING DATE: 2000-10-02
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3382
; LENGTH: 3062
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U51333
US-09-880-107-3382

Query Match 79.2%; Score 19; DB 9; Length 3062;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGACC 19
Db 1081 CAGCCACCAAGAGACC 1063

RESULT 7
US-10-342-887-728/c
; Sequence 728, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yundong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR FILING DATE: 2003-01-15
; PRIOR FILING DATE: 2001-06-18
; PRIOR FILING DATE: 2002-05-14
; PRIOR FILING DATE: 2002-05-14
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 728
; LENGTH: 3062
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-728

Query Match 79.2%; Score 19; DB 13; Length 3062;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGACC 19
Db 1081 CAGCCACCAAGAGACC 1063

RESULT 8
US-10-172-118-728/c
; Sequence 728, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yundong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR FILING DATE: 2002-05-14
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 728
; LENGTH: 3062
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 002115
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-728

Query Match 79.2%; Score 19; DB 13; Length 3062;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGACC 19
Db 1081 CAGCCACCAAGAGACC 1063

RESULT 9
US-10-717-597-214/c
; Sequence 214, Application US/10717597
; Publication No. US20040110221A1
; GENERAL INFORMATION:
; APPLICANT: Wylech
; APPLICANT: Burczynski, Michael E.
; APPLICANT: Twine, Natalie C.
; APPLICANT: Dornier, Andrew J.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Stonim, Donna K.
; APPLICANT: Scover, Jennifer A.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
; FILE REFERENCE: AM101080L
; CURRENT APPLICATION NUMBER: US/10/717,597
; CURRENT FILING DATE: 2003-11-21
; PRIOR FILING DATE: 2003-04-03
; PRIOR FILING DATE: 2003-04-03
; PRIOR FILING DATE: 2002-11-21
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 214
; LENGTH: 3062
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-717-597-214

Query Match 79.2%; Score 19; DB 17; Length 3062;

Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
|||||

Db 1081 CAGCCACCAAGAGAGACC 1063

RESULT 10
US-10-177-809-1/c

/ Sequence 1, Application US/10177809
/ Publication No. US20030148305A1
/ GENERAL INFORMATION:
/ APPLICANT: Grose, William M.,
/ APPLICANT: Alsobrook, John P.,
/ APPLICANT: Lepley, Denise M.,
/ APPLICANT: Burgess, Catherine E.,
/ APPLICANT: Bader, Joel S.,
/ APPLICANT: Banaal, Aruna,
/ APPLICANT: Pena, Carol E.A.,
/ APPLICANT: Shinkets, Richard A.,
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Anderson, David
/ APPLICANT: Zhong, Mei
/ APPLICANT: Miller, Charles
/ APPLICANT: Vernet, Corine
/ APPLICANT: Hjalte, Tor
/ TITLE OF INVENTION: NUCLEIC ACIDS, POLYPEPTIDES, SINGLE NUCLEOTIDE POLYMORPHISMS AND
/ FILE REFERENCE: 21402-390
/ CURRENT APPLICATION NUMBER: US/10/177,809
/ PRIOR FILING DATE: 2003-01-28
/ PRIOR APPLICATION NUMBER: 60/311,285
/ PRIOR FILING DATE: 2001-08-09
/ PRIOR APPLICATION NUMBER: 60/299,949
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: 60/300,290
/ PRIOR FILING DATE: 2001-06-22
/ PRIOR APPLICATION NUMBER: 60/302,883
/ PRIOR FILING DATE: 2001-07-03
/ PRIOR APPLICATION NUMBER: 60/327,345
/ PRIOR FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 60/327,467
/ PRIOR FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 60/327,892
/ PRIOR FILING DATE: 2001-10-09
/ PRIOR APPLICATION NUMBER: 60/354,586
/ PRIOR FILING DATE: 2002-02-05
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: Custom
/ SEQ ID NO 1
/ LENGTH: 3205
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (75)..(1164)
US-10-177-809-1

Query Match 79.2%; Score 19; DB 15; Length 3205;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
|||||

Db 1081 CAGCCACCAAGAGAGACC 1063

RESULT 11
US-10-177-809-3/c

/ Sequence 3, Application US/10177809

/ Publication No. US20030148305A1
/ GENERAL INFORMATION:
/ APPLICANT: Grose, William M.,
/ APPLICANT: Alsobrook, John P.,
/ APPLICANT: Lepley, Denise M.,
/ APPLICANT: Burgess, Catherine E.,
/ APPLICANT: Bader, Joel S.,
/ APPLICANT: Banaal, Aruna,
/ APPLICANT: Pena, Carol E.A.,
/ APPLICANT: Shinkets, Richard A.,
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Anderson, David
/ APPLICANT: Zhong, Mei
/ APPLICANT: Miller, Charles
/ APPLICANT: Vernet, Corine
/ APPLICANT: Hjalte, Tor
/ TITLE OF INVENTION: NUCLEIC ACIDS, POLYPEPTIDES, SINGLE NUCLEOTIDE POLYMORPHISMS AND
/ FILE REFERENCE: 21402-390
/ CURRENT APPLICATION NUMBER: US/10/177,809
/ PRIOR FILING DATE: 2003-01-28
/ PRIOR APPLICATION NUMBER: 60/311,285
/ PRIOR FILING DATE: 2001-08-09
/ PRIOR APPLICATION NUMBER: 60/299,949
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: 60/300,290
/ PRIOR FILING DATE: 2001-06-22
/ PRIOR APPLICATION NUMBER: 60/302,883
/ PRIOR FILING DATE: 2001-07-03
/ PRIOR APPLICATION NUMBER: 60/327,345
/ PRIOR FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 60/327,467
/ PRIOR FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 60/327,892
/ PRIOR FILING DATE: 2001-10-09
/ PRIOR APPLICATION NUMBER: 60/354,586
/ PRIOR FILING DATE: 2002-02-05
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: Custom
/ SEQ ID NO 3
/ LENGTH: 3205
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-177-809-3

Query Match 79.2%; Score 19; DB 15; Length 3205;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
|||||

Db 1081 CAGCCACCAAGAGAGACC 1063

RESULT 12
US-10-177-809-5/c

/ Sequence 5, Application US/10177809
/ Publication No. US20030148305A1
/ GENERAL INFORMATION:
/ APPLICANT: Grose, William M.,
/ APPLICANT: Alsobrook, John P.,
/ APPLICANT: Lepley, Denise M.,
/ APPLICANT: Burgess, Catherine E.,
/ APPLICANT: Bader, Joel S.,
/ APPLICANT: Banaal, Aruna,
/ APPLICANT: Pena, Carol E.A.,
/ APPLICANT: Shinkets, Richard A.,
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Anderson, David

APPLICANT: Zhong, Mei
APPLICANT: Miller, Charles
APPLICANT: Vermet, Corine
APPLICANT: Hjalte, Tord
TITLE OF INVENTION: NUCLEIC ACIDS, POLYPEPTIDES, SINGLE NUCLEOTIDE POLYMORPHISMS AND
FILE REFERENCE: 21402-390
CURRENT APPLICATION NUMBER: US/10/177,809
CURRENT FILING DATE: 2003-01-28
PRIOR APPLICATION NUMBER: 60/311,285
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/299,949
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: 60/300,290
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: 60/302,883
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: 60/327,345
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 60/327,467
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 60/327,892
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/354,586
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Custom
SEQ ID NO 5
LENGTH: 3205
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-809-5

Query Match 79.2%; Score 19; DB 15; Length 3205;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
DB 1081 CAGCCACCAAGAGAGACC 1063

RESULT 13
US-10-152-319A-1767/c
Sequence 1767, Application US/10152319A
Publication No. US20040072160A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: HIGGS, Brandon
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5089-US
CURRENT APPLICATION NUMBER: US/10/152,319A
CURRENT FILING DATE: 2002-05-22
PRIOR APPLICATION NUMBER: US 60/292,335
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/297,523
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,925
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,810
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/303,807
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/303,808
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/315,047
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: US 60/324,928
PRIOR FILING DATE: 2001-09-27

PRIOR APPLICATION NUMBER: US 60/330,867
PRIOR FILING DATE: 2001-11-01
PRIOR APPLICATION NUMBER: US 60/330,462
PRIOR FILING DATE: 2001-10-22
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 2221
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1767
LENGTH: 3692
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. NM_022179
US-10-152-319A-1767

Query Match 79.2%; Score 19; DB 12; Length 3692;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
DB 1087 CAGCCACCAAGAGAGACC 1069

RESULT 14
US-10-191-803-138/c
Sequence 138, Application US/10191803
Publication No. US20040014040A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: JOHNSON, Kory
APPLICANT: HIGGS, Brandon
APPLICANT: CASTLE, Arthur
APPLICANT: ELASHOFF, Michael
TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
FILE REFERENCE: 44921-5090US
CURRENT APPLICATION NUMBER: US/10/191,803
CURRENT FILING DATE: 2002-07-10
PRIOR APPLICATION NUMBER: US 60/303,819
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/305,623
PRIOR FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: US 60/369,351
PRIOR FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: US 60/377,611
PRIOR FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 1140
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 138
LENGTH: 3692
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20040014040A1 NM_022179
US-10-191-803-138

Query Match 79.2%; Score 19; DB 16; Length 3692;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGACC 19
DB 1087 CAGCCACCAAGAGAGACC 1069

RESULT 15
US-09-764-861-63/c
Sequence 63, Application US/09764861
Publication No. US20030171252A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

; FILE REFERENCE: PT204
; CURRENT APPLICATION NUMBER: US/09/764,861
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 63
; LENGTH: 18772
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (9782)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-861-63

```

```

Query Match          79.2%; Score 19; DB 10; Length 18772;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CAGCCACCAAGAGAGGCC 19
Db      708 CAGCCACCAAGAGAGGCC 690

```

Search completed: July 30, 2004, 22:13:04
Job time : 410 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 20:13:41 ; Search time 2963 Seconds
(without alignments)
241.881 Million cell updates/sec

Title: US-09-831-272-7
Sequence: 1 cagccaccaagagagaccagat 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_escba:*
2: em_escbm:*
3: em_escbn:*
4: em_escbu:*
5: em_escbv:*
6: em_escbl:*
7: em_escro:*
8: em_escr:*
9: gb_esc1:*
10: gb_esc2:*
11: gb_esc3:*
12: gb_esc4:*
13: gb_esc5:*
14: gb_esc6:*
15: em_escfun:*
16: em_escfm:*
17: em_escfh:*
18: em_escfi:*
19: em_escfl:*
20: em_escfv:*
21: em_escfw:*
22: em_escfx:*
23: em_escfy:*
24: em_escfz:*
25: em_escga:*
26: em_escgb:*
27: em_escgc:*
28: gb_esc1:*
29: gb_esc2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	85.0	934	14	CK269862 EST715940
2	19.8	82.5	347	29	CC468543 CH240.139
3	19.8	82.5	528	9	AV529210 AV529210
4	19.8	82.5	531	28	A2114521 RPT-23-4

Result No.	Score	Query Match	Length	ID	Description
5	19.8	82.5	534	9	AV529714 AV529714
6	19.8	82.5	705	14	CK317567 CK317567
7	19.8	82.5	1076	10	EG024130 EG024130
8	19.8	82.5	1082	10	BE729686 BE729686
9	19.2	80.0	242	9	AV271379 AV271379
10	19.2	80.0	257	9	AV333896 AV333896
11	19.2	80.0	708	13	BU026771 BU026771
12	19.2	80.0	728	28	AZ312990 AZ312990
13	19.2	80.0	899	29	CC518089 CC518089
14	19.2	80.0	359	10	BF901013 BF901013
15	19.2	80.0	897	14	CD557470 CD557470
16	18.8	78.3	321	14	CB972921 CB972921
17	18.8	78.3	452	14	CD802324 CD802324
18	18.8	78.3	463	13	BF427213 BF427213
19	18.8	78.3	466	13	BQ798315 BQ798315
20	18.8	78.3	506	10	AW140945 AW140945
21	18.8	78.3	510	12	BG813094 BG813094
22	18.8	78.3	550	9	AA990846 AA990846
23	18.8	78.3	562	10	AW913869 AW913869
24	18.8	78.3	566	28	AQ962724 AQ962724
25	18.8	78.3	591	14	CF413809 CF413809
26	18.8	78.3	622	14	CA352741 CA352741
27	18.8	78.3	635	10	BB625968 BB625968
28	18.8	78.3	643	13	BQ208827 BQ208827
29	18.8	78.3	653	13	BQ193103 BQ193103
30	18.8	78.3	656	10	BE270340 BE270340
31	18.8	78.3	669	28	AQ962725 AQ962725
32	18.8	78.3	670	13	BQ799382 BQ799382
33	18.8	78.3	670	14	CF384461 CF384461
34	18.8	78.3	704	14	CA505029 CA505029
35	18.8	78.3	706	13	BX845101 BX845101
36	18.8	78.3	710	28	CC157138 CC157138
37	18.8	78.3	719	10	BF966746 BF966746
38	18.8	78.3	759	14	CB319952 CB319952
39	18.8	78.3	796	13	BQ799340 BQ799340
40	18.8	78.3	830	12	BG518541 BG518541
41	18.8	78.3	832	13	BU428105 BU428105
42	18.8	78.3	845	29	CG820963 CG820963
43	18.8	78.3	892	13	BU910667 BU910667
44	18.8	78.3	913	29	CG054407 CG054407
45	18.8	78.3	913	29	CG054407 CG054407

ALIGNMENTS

RESULT 1
CK269862/c 934 bp mRNA linear EST 12-DEC-2003
LOCUS EST715940 potato abiotic stress cDNA library Solanum tuberosum cDNA
DEFINITION clone POACP48 5' end, mRNA sequence.
ACCESSION CK269862
VERSION CK269862.1 GI:39826840
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 934)
AUTHORS Buell, C.R., Hart, A., Zismann, V., Karanycheva, S.A. and Baker, B.
TITLE Generation of ESTs from abiotic stressed potato tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-arrays@tigr.org
Clones can be requested from TIGR via potato@tigr.org
Seq primer: ATP TAG GTG ACA CTA TAG.
Location/Qualifiers
1..934
/organism="Solanum tuberosum"
/mol_type="mRNA"

FEATURES

source

```

/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POACp48"
/tissue_type="abiotic stress treated leaf and root tissue"
/lab_host="DH10B-TonA"
/clone_idb="potato abiotic stress cDNA library"
/notes="Vector: PCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: Solanum tuberosum var. Kennebec plants were
grown from cuttings on a 16hr light/8 hr dark cycle at 25
C for 3-4 weeks. Abiotic stress conditions were applied to
four separate sets of plants. Set 1 involved saturation of
the soil with 150 mM NaCl and tissues were harvested at
following application of the salt stress (leaves: 2hr,
6hr, 12hr, 1d, 2d, and 4d; roots: 2hr, 6hr, 12hr, 1d,
and 2d). Set 2 were grown under the standard conditions and then
were water stressed by withdrawal of further watering
applications. Drought stressed plants were harvested after
cessation of watering (leaves: 3d, 5d, and 7d; roots: 3d
and 5d). Set 3 were grown under the standard conditions
and then were cold stressed by placement at 4 C. Cold
stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d,
and 4d. Set 4 were grown under the standard conditions and
then were heat stressed by placement at 35 C. Heat
stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
2 d, 1 d, and heat-stressed roots were harvested at 6 hr,
12 hr, 1 d, and 4d. RNA was isolated from all tissues and
equal RNA from each tissue and stress was pooled to
construct the cDNA library. RNA sample."

```

ORIGIN

```

Query Match      85.0%; Score 20.4; DB 14; Length 934;
Best Local Similarity 95.5%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 2 AGCCACCAAGAGAGACCCAGAA 23
    |||||||
Db 171 AGCCACCAAGAGAGACCCAGAA 150

```

```

RESULT 2
CC468543/c      CC468543      347 bp      DNA      linear      GSS 12-JUN-2003
LOCUS           CH240_139N06.TV CHORI-240 Bos taurus genomic clone CH240_139N06,
DEFINITION      genomic survey sequence.
ACCESSION       CC468543
VERSION         CC468543.1 GI:31654775
KEYWORDS        GSS.
SOURCE          Bos taurus (cow)
ORGANISM        Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 347)
Coستا,J.N., Mota,M. and Caetano,A.R.
Brazil's Contribution to End-Sequencing the Bovine BAC Library
CHORI-240
Unpublished (2003)
Other_GSSs: CH240_139N06.TJ
Contact: Caetano AR
Department of Biotechnology
Embrapa Recursos Geneticos e Biotecnologia
Parque Estacao Biologica, Final Av. W/5 Norte, Brasilia-DF C.P.
02372 70770-900 Brasil
Tel: 55 61 448 4778
Fax: 55 61 340 3658
Email: acetano@cenargen.embrapa.br
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). Bases shown have Phred
quality value equal to or higher than 20. Bases with quality value
below 20 were masked with 'N'. For BAC library availability, please
contact Pieter de Jong (pdejong@mail.cho.org). Clones may be
purchased from BACPAC Resources

```

```

(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by Embrapa Recursos Geneticos e Biotecnologia
with financing from Conselho Nacional de Desenvolvimento Cientifico
e Tecnológico (CNPq), Brazil.
Plate: 139 row: N column: 06
Seq primer: T7
Class: BAC ends
High quality sequence stop: 347.
Location/Qualifiers
1..347
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Bred: Hereford"
/db_xref="taxon:9913"
/clone="CH240_139N06"
/sex="Male"
/cell_type="Blood"
/clone_idb="CHORI-240"
/notes="Vector: PARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domingo 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

```

ORIGIN

```

Query Match      82.5%; Score 19.8; DB 29; Length 347;
Best Local Similarity 91.3%; Pred. No. 6.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 2 AGCCACCAAGAGAGACCCAGAA 24
    |||||||
Db 298 AGCCACCAAGAGAGACCCAGAA 276

```

```

RESULT 3
AV529210/c      AV529210      528 bp      mRNA      linear      EST 01-SEP-2000
LOCUS           AV529210 Arabidopsis thaliana aboveground organs two to six-week
DEFINITION      old Arabidopsis thaliana cDNA clone APZL30407R 5', mRNA sequence.
ACCESSION       AV529210
VERSION         AV529210.1 GI:8689493
KEYWORDS        EST.
SOURCE          Arabidopsis thaliana (thale cress)
ORGANISM        Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 528)
Asamizu,B., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
Yana 1532-3, Kisarazu, Chiba 292-0612, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..528
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APZL30407R"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_idb="Arabidopsis thaliana aboveground organs two to
six-week old"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI"

```

FEATURES

source

ORIGIN

Query Match 82.5%; Score 19.8; DB 9; Length 528;
 Best Local Similarity 91.3%; Pred. No. 7.2e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGACCCAGAA 24
 |||||
 44 AACCCACCAAGAGACCCAAAT 22

RESULT 4
 AZ114521 531 bp DNA linear GSS 12-MAY-2000
 LOCUS RPCI-23-460E1.TV RPCI-23 Mus musculus genomic clone RPCI-23-460E1,
 DEFINITION genomic survey sequence.
 ACCESSION AZ114521
 VERSION AZ114521.1 GI:7774836
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 1 (bases 1 to 531)
 Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S.,
 Akinet, B., Levins, M., McGann, S., Tsengaye, G., Geer, K., Krol, M., de
 Jong, P. and Fraser, C. M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-460E1.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 460 row: E column: 1
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..531
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-460E1"
 /sex="Female"
 /lab_host="DH10B"
 /clone_1lb="RPCI-23"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI MethyIase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 82.5%; Score 19.8; DB 28; Length 531;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGACCCAGAA 23
 |||||
 150 CAGCCACCAAGAGACCCAGAA 172

RESULT 5
 AV529714/c 534 bp mRNA linear EST 01-SEP-2000
 LOCUS AV529714 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION old Arabidopsis thaliana CDNA clone APZ44905' 5', mRNA sequence.

ACCESSION AV529714
 VERSION AV529714.1 GI:8689997
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS Asanizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7 (3), 175-180 (2000)
 MEDLINE 20363093
 PUBMED 10907847
 COMMENT Contact: Erika Asanizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asanizu@kazusa.or.jp URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES

Location/Qualifiers
 1..534
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="APZ44905R"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /clone_1lb="Arabidopsis thaliana aboveground organs two to
 six-week old"
 /note="vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN

Query Match 82.5%; Score 19.8; DB 9; Length 534;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCCACCAAGAGACCCAGAA 24
 |||||
 Db 229 AACCCACCAAGAGACCCAAAT 207

RESULT 6
 CK317567/c 705 bp mRNA linear EST 16-DEC-2003
 LOCUS CK317567 VRJ376 3', mRNA sequence.
 DEFINITION VRJ376 3', mRNA sequence.

ACCESSION CK317567
 VERSION CK317567.1 GI:39952381
 KEYWORDS EST.
 SOURCE Vitis riparia
 ORGANISM Vitis riparia
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.
 1 (bases 1 to 705)
 Fennell, A. and Mathiason, K.
 Expressed sequence tags from endodermant Vitis riparia buds
 Unpublished (2003)
 COMMENT Contact: Anne Fennell
 Horticulture, Forestry, Landscape and Parks Department
 South Dakota State University
 Northern Plains Biostress Laboratory, Box 2140A, Brookings, SD
 57007-0996, USA
 Tel: 605 688 6373

Fax: 605 688 4713
Email: Anne.Fennell@edstate.edu
Seq primer: polyT.

FEATURES

Location/Qualifiers

1..705

/organism="Vitis riparia"

/mol_type="mRNA"

/db_xref="taxon:96939"

/clone="VRJ376"

/dev_stage="dormant"

/lab_host="DH10B"

/clone_lib="Vitis riparia endoderm bud - VRJ"

/note="Organ: bud; Vector: pSPORT 1; Site 1: SalI; Site 2:

NotI; VRJ is a cDNA library of Vitis riparia endoderm

buds. Endoderm buds were collected from insect and

disease free vines that were induced into dormancy with

short photoperiods. The directionally oriented library was

constructed according to Gibco BRL Superscript Plasmid

System for cDNA synthesis and plasmid cloning."

ORIGIN

Query Match 82.5%; Score 19.8; DB 14; Length 705;
Best Local Similarity 91.3%; Pred. No. 7.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGCCACCAAGAGAGACCCGAAT 24

DB 297 AGCCACCAAGAGAGACCCGAAT 275

RESULT 7 1076 bp mRNA linear EST 24-JAN-2001
LOCUS BG024130
DEFINITION 602303003P1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4394473 5',
mRNA sequence.

ACCESSION BG024130

VERSION BG024130.1 GI:12409388

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 1076)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LILAM10090 row: e column: 02

High quality sequence stop: 610.

FEATURES

source

1..1076

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4394473"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_88"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: SalI; Cloned unidirectionally;

oligo-dt primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: This is a NIH_MGC Library."

ORIGIN

Query Match 82.5%; Score 19.8; DB 10; Length 1076;
Best Local Similarity 91.3%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAGCCACCAAGAGAGACCCGAAT 23

DB 905 CAGCCACCAAGAGAGAGACCCGAAT 927

RESULT 8 1082 bp mRNA linear EST 15-SEP-2000
LOCUS BE729686
DEFINITION 60156283F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3832425 5',
mRNA sequence.

ACCESSION BE729686

VERSION BE729686.1 GI:10143678

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 1082)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov

Plate: LILCM509 row: j column: 10

High quality sequence stop: 737.

FEATURES

source

1..1082

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3832425"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_20"

/note="Organ: skin; Vector: pORF7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 82.5%; Score 19.8; DB 10; Length 1082;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGCCACCAAGAGAGACCCGAAT 24

DB 834 AGCCACCAAGAGAGAGACCCGAAT 856

RESULT 9 242 bp mRNA linear EST 05-NOV-1999
LOCUS AV271379
DEFINITION AV271379 RIKEN full-length enriched, adult male testis (DH10B) Mus
musculus cDNA clone 4931408J02 3', mRNA sequence.

ACCESSION AV271379

VERSION AV271379.1 GI:6259416

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

ORIGIN

REFERENCE AUTHORS

REFERENCE	AUTHORS
1 (bases 1 to 242)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Komono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukuishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Iwawa, M., Kadoori, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Shibata, K., Shibaie, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomnaga, N., Tsunoda, Y., Wataniki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	
RIKEN Mouse ESTs (Komono, H., et al. 1999)	
Unpublished (1999)	
CONTACT:	Yoshihide Hayashizaki

FEATURES
SOL

1. .242
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4931408J02"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
GAGGAGAGAGAGATCCACAGCGCTTTTTTTTTTTTTTNN 3'). cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'
GAGGAGAGATTCGCGATTAAATTAATTAATCCCCCCCCC 3'. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda VEC 1. Cloning sites, 5' end: SalI; 3' end: BamHI."

[illegible]

```

OY      1 CAGCCACCAAGAAGAGACCCAGAAAT 24
          ||||| | ||||| ||||| |||||
DB      38 CAGCTATCAAGAAGAGACCCAGATT 61

```

RESULT 10	
AV333896	
LOCUS	257 bp mRNA linear EST 11-NOV-1999
DEFINITION	AV333896 RIKEN full-length enriched, adult male medulla oblongata
ACCESSION	Mus musculus cDNA clone 633052C10 3', mRNA sequence.
VERSION	AV333896
KEYWORDS	AV333896.1 GI:6373948
SOURCE	EST.
	Mus musculus (house mouse)

FEATURES
SOURCE

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330552C10"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="PH108"
/clone_1ib="RIKEN full-length enriched, adult male medulla oblongata"
/notes="Site_1: Sall; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

```

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGATCCAGACCTTTTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rct = 10.0 and subtraction to Rct = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATCTCGATTAATTATTAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 80.0%; Score 19.2; DB 9; Length 257;
Best Local Similarity 87.5%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 CAGCCACCAAGAGACCCAGAT 24
|||||
54 CAGCCATCAGAGACCCAGATT 77

RESULT 11
BU026771 708 bp mRNA linear EST 23-AUG-2002
LOCUS OHG17P22.yg.ab1 OH ERGHJ sunflower RHA280 Helianthus annuus cDNA
DEFINITION clone OHG17P22, mRNA sequence.
ACCESSION BU026771
VERSION BU026771.1 GI:22462291
KEYWORDS EST.
SOURCE Helianthus annuus (common sunflower)
ORGANISM Helianthus annuus

REFERENCE
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositeae Genome Project
http://comgenomics.ucdavis.edu/
Unpublished (2002)

JOURNAL
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Assumundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
Singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: OHG17 row: P column: 22.

FEATURES
source

1..708
Location/Qualifiers
/organism="Helianthus annuus"
/mol_type="mRNA"
/cultivar="RHA280"
/db_xref="taxon:4232"
/clone="OHG17P22"
/lab_host="E.coli"
/clone_lib="OH_ERGHJ sunflower RHA280"
/note="Vector: pBRCNDS1AB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/

ORIGIN

Query Match 80.0%; Score 19.2; DB 13; Length 708;
Best Local Similarity 87.5%; Pred. No. 1.4e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 CAGCCACCAAGAGACCCAGAT 24
|||||
147 CAGCCACCAAGAGACCCAGATT 170

RESULT 12
AZ312990/c 728 bp DNA linear GSS 29-SEP-2000
LOCUS IM0023H08F Mouse 10kb plasmid U06C1M library Mus musculus genomic
DEFINITION clone U06C1M0023H08 F, genomic survey sequence.
ACCESSION AZ312990
VERSION AZ312990.1 GI:10357470
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 728)
Dunn, P., Aoyagi, A., Barber, M., Baecorn, T., Duval, B., Hamill, C., Isiam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)

JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0029 row: H column: 08
Seq primer: CATTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 728.

FEATURES
source

1..728
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0023H08"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to

ORIGIN

adaptoed vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 80.0%; Score 19.2; DB 28; Length 728;
Best Local Similarity 87.5%; Pred. No. 1.4e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGAGAGAT 24
58 CAGACACCAATGACACCAAGAT 35

RESULT 13

CC518089 899 bp DNA linear GSS 17-JUN-2003
LOCUS CH240_364J21.T7 CHORI-240 Bos taurus genomic clone CH240_364J21,
DEFINITION genomic survey sequence.

ACCESSION CC518089
VERSION CC518089.1 GI:31836377
KEYWORDS GSS.

SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE

AUTHORS

1 (bases 1 to 899)
Holt, R., Scott, J., Yang, G., Barber, S., Small, D., Prabhu, A.-L.,
Tsal, M., Cloutier, A., Lee, D., Gilm, N., Olson, T., Mayo, M., Chiu, R.,
Butterfield, Y., Kirkpatrick, R., Liu, J., Gilm, R., Chan, A., Jones, S.,
Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,
Schein, J., Marra, M., de Jong, P., McWilliam, S., Barris, W.,
Dalympole, B. P. and Teilmann, R.
Bovine BAC End Sequences from Library CHORI-240, PLATTS 294 to 398
Unpublished (2003)
Other GSSs: CH240_364J21.TARBACT13P2
Contact: Rob Holt

JOURNAL

COMMENT

Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bccsc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by CSIRO Livestock Industries, Australia and the
British Columbia Genome Sciences Centre, Canada.
Place: 364 row: J column: 21
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..899
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Bred: Hereford"
/db_xref="taxon:9913"
/clone="CH240_364J21"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: PTARBAC1.3; Site 1: Mbol; Site 2: Mbol;
Hereford bull LI Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 80.0%; Score 19.2; DB 29; Length 899;
Best Local Similarity 87.5%; Pred. No. 1.5e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGAGAGAT 24
598 CAGCCACCAATGAGAGAGAGAT 621

RESULT 14

BF901013/c 359 bp mRNA linear EST 18-JAN-2001
LOCUS PM1-MT0203-081200-003-b03 MT0203 Homo sapiens cDNA, mRNA sequence.

DEFINITION BF901013
VERSION BF901013.1 GI:12292472
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 359)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V.,
O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

JOURNAL

COMMENT

This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM1&ct=PM1-MT0203-
081200-003-b03&ct3=2000-12-08&ct4=1)
Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 359.
Location/Qualifiers
1..359
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0203"
/note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

FEATURES

source

Location/Qualifiers
1..359
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0203"
/note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

ORIGIN

Query Match 79.2%; Score 19; DB 10; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCCACCAAGAGAGAGAT 19
137 CAGCCACCAAGAGAGAGAT 119

RESULT 15

CD557470/c 897 bp mRNA linear EST 11-JUN-2003
LOCUS CD557470
DEFINITION AGENCOURT_14413595 NIH_MGC_180 Homo sapiens cDNA clone

IMAGE:30387677 5', mRNA sequence.

ACCESSION CD557470

VERSION CD557470.1 GI:31583538

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS 1 (bases 1 to 897)

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1995)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: c9apba-remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: NDAM458 row: F column: 06

High quality sequence stop: 627.

Location/Qualifiers

1. 897

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:30387677"

/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"

/clone_lib="NIH_MGC_180"

/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;

Site_2: EcoRV (destroyed); Library is oligo-dT primed and

directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.68 kb. Library was

constructed by (Invitrogen). Note: this is a NIH_MGC

library."

ORIGIN

Query Match 79.2%; Score 19; DB 14; Length 897;

Best Local Similarity 100.0%; Pred. No. 1.8e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0;

Gy 1 CAGCCACCAAGAGAGACCC 19

Db 417 CAGCCACCAAGAGAGACCC 399

Search completed: July 30, 2004, 22:04:43

Job time : 2973 secs